

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 18, 2005, 15:30:38 ; Search time 139 Seconds

(without alignments)
795.737 Million cell updates/sec

Title: US-10-791-017A-2_COPY_319_656

Perfect score: 1922
Sequence: 1 GGRGGMGSGRGGRRGPGG.....GGPRKDKGRRHRRDRPY 338Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1380268 seqs, 327241040 residues

Total number of hits satisfying chosen parameters: 1380268

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

Published Applications AA:*
1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep:*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep:*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep:*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep:*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/2/pubpaa/US10D_PUBCOMB.pep:*
17: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
18: /cgn2_6/ptodata/2/pubpaa/US11_NEW_PUB.pep:*
19: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
20: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1922	100.0	600	US-10-094-749-2713	Sequence 2713, App1
2	1922	100.0	656	US-10-791-017A-2	Sequence 2, App1
3	878	45.7	525	US-10-755-889-650	Sequence 650, App
4	799.5	41.6	260	US-10-408-765A-116	Sequence 116, App
5	623	32.4	156	US-09-925-301-1511	Sequence 1511, App
6	475	22.7	83	US-09-864-761-42397	Sequence 42397, A
7	423	22.0	467	US-10-437-963-121926	Sequence 121926, A
8	420	21.9	395	US-10-425-114-69744	Sequence 69744, A
9	319.5	16.6	578	US-10-437-963-164510	Sequence 164510, A
10	303	15.8	197	US-10-424-599-167688	Sequence 167688, A
11	303	15.8	1078	US-10-058-124-21	Sequence 21, App1
12	303	15.8	1466	US-10-402-089-4	Sequence 4, App1
13	303	15.8	1466	US-10-402-089-6	Sequence 6, App1

14	303	15.8	1466	US-10-402-072A-4	Sequence 4, App1
15	303	15.8	1466	US-10-402-072A-6	Sequence 6, App1
16	301	15.7	1466	US-09-918-715-226	Sequence 226, App
17	301	15.7	1466	US-10-177-293-68	Sequence 68, App1
18	301	15.7	1466	US-10-301-822-33	Sequence 33, App1
19	301	15.7	1466	US-10-257-021-72	Sequence 72, App1
20	301	15.7	1466	US-10-357-851-3	Sequence 3, App1
21	301	15.7	1466	US-10-358-024-3	Sequence 3, App1
22	301	15.7	1466	US-10-734-564-103	Sequence 103, App
23	300.5	15.6	1466	US-10-402-089-12	Sequence 12, App1
24	300.5	15.6	1466	US-10-402-072A-12	Sequence 12, App1
25	284.5	14.8	1462	US-10-437-963-120427	Sequence 120427, A
26	275.5	14.3	1496	US-10-177-293-70	Sequence 70, App1
27	275.5	14.3	1496	US-10-301-822-35	Sequence 35, App1
28	275.5	14.3	1496	US-10-236-031B-74	Sequence 74, App1
29	275.5	14.3	1496	US-10-468-091-22	Sequence 22, App1
30	275.5	14.3	1496	US-10-788-792-248	Sequence 248, App
31	272.5	14.2	651	US-10-488-056-45	Sequence 45, App1
32	272.5	14.2	1449	US-10-402-089-8	Sequence 8, App1
33	272.5	14.2	1449	US-10-402-072A-8	Sequence 8, App1
34	271.5	14.1	968	US-10-291-172-739	Sequence 739, App
35	271.5	14.1	968	US-10-221-278-739	Sequence 739, App
36	270.5	14.1	1046	US-10-156-761-10088	Sequence 10088, A
37	270	14.0	1487	US-10-468-091-6	Sequence 6, App1
38	270	14.0	1497	US-10-468-091-23	Sequence 23, App1
39	268.5	14.0	386	US-10-408-765A-959	Sequence 959, App
40	268.5	14.0	1463	US-10-402-089-2	Sequence 2, App1
41	268.5	14.0	1463	US-10-402-072A-2	Sequence 2, App1
42	267	13.9	396	US-10-424-599-251915	Sequence 251915, A
43	266.5	13.9	940	US-10-291-172-363	Sequence 363, App
44	266.5	13.9	940	US-10-221-278-363	Sequence 363, App
45	265.5	13.8	1453	US-10-468-091-26	Sequence 26, App1

ALIGNMENTS

RESULT 1
US-10-094-749-2713
; Sequence 2713, Application US/10094749
; Publication No. US20030219741A1
GENERAL INFORMATION:
; APPLICANT: ISOGAI, TAKAO
; APPLICANT: SUGIYAMA, TOMOYASU
; APPLICANT: OTSUKI, TETSUJI
; APPLICANT: WAKAMATSU, AI
; APPLICANT: SATO, HIROYUKI
; APPLICANT: ISHII, SHIZUKO
; APPLICANT: YAMAMOTO, JUN-ICHI
; APPLICANT: ISONO, YUUKO
; APPLICANT: HIO, YURI
; APPLICANT: OTSUKA, KAORU
; APPLICANT: NAGAI, KEIICHI
; APPLICANT: IRIE, RYOTARO
; APPLICANT: TAMECHIKA, ICHIRO
; APPLICANT: SEKI, NAOHITO
; APPLICANT: YOSHIKAWA, TSUTOMU
; APPLICANT: OTSUKA, MOTOTYUKI
; APPLICANT: NAGAHARA, KENJI
; APPLICANT: MASUHO, YASUHIKO
TITLE OF INVENTION: NOVEL FULL-LENGTH CDNA
; FILE REFERENCE: 084335/0160
; CURRENT APPLICATION NUMBER: US/10/094, 749
; PRIOR FILING DATE: 2002-03-12
; PRIOR APPLICATION NUMBER: 60/350, 435
; PRIOR FILING DATE: 2002-01-24
; PRIOR APPLICATION NUMBER: JP 2001-328381
; NUMBER OF SEQ ID NOS: 3381
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 2713
; LENGTH: 600
; TYPE: PRT

ORGANISM: Homo sapiens
US-10-094-749-2713

Query Match 100.0%; Score 1922; DB 15; Length 600;
Best Local Similarity 100.0%; Pred. No. 3, 1e-128;
Matches 338; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 GGRGMSAGERGPFKPGPMDEGPDLDLGPVDDPDESDNSAIYVGLNDSVTLLDILA 60
263 GGRGMSAGERGPFKPGPMDEGPDLDLGPVDDPDESDNSAIYVGLNDSVTLLDILA 322
61 DFFKQGVVNMNRKTQPMIHYLDKETGPKGDAATVSYEDPTAKAAVEMFGKDFQGS 120
323 DFFKQGVVNMNRKTQPMIHYLDKETGPKGDAATVSYEDPTAKAAVEMFGKDFQGS 382
121 KLVSLARKKPPMNSMRGGLPPREGGMPPLRLGPGGPGGPGMGRGGRGDRGFP 180
383 KLVSLARKKPPMNSMRGGLPPREGGMPPLRLGPGGPGGPGMGRGGRGDRGFP 442
181 PRGPRGSRGNPSGGVNVQHRAGDMQCPNPGCGNQNFMARTECNQCAKPKREGFLPPFP 240
443 PRGPRGSRGNPSGGVNVQHRAGDMQCPNPGCGNQNFMARTECNQCAKPKREGFLPPFP 502
241 PGDGRGPGGPGMGRGGRGGLMDRGPGCMFRGGRGGRGGRGGRGGRGGRG 300
503 PGDGRGPGGPGMGRGGRGGLMDRGPGCMFRGGRGGRGGRGGRGGRGGRG 562
301 GGPPLMEQMGRRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRG 338
563 GGPPLMEQMGRRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRG 600

RESULT 2

US-10-791-017a-2
Sequence 2, Application US/10791017A
Publication No. US20040197827A1
GENERAL INFORMATION:
APPLICANT: JENAPHARM GmbH & Co. KG
TITLE OF INVENTION: Methods for Determining Hormonal Effects of Substances
FILE REFERENCE: Pat 3684/11
CURRENT APPLICATION NUMBER: US/10/791, 017A
NUMBER OF SEQ ID NOS: 7
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 2
LENGTH: 656
TYPE: PRT
ORGANISM: Homo sapiens
US-10-791-017a-2

Query Match 100.0%; Score 1922; DB 16; Length 656;
Best Local Similarity 100.0%; Pred. No. 3, 3e-128;
Matches 338; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 GGRGMSAGERGPFKPGPMDEGPDLDLGPVDDPDESDNSAIYVGLNDSVTLLDILA 60
319 GGRGMSAGERGPFKPGPMDEGPDLDLGPVDDPDESDNSAIYVGLNDSVTLLDILA 378
61 DFFKQGVVNMNRKTQPMIHYLDKETGPKGDAATVSYEDPTAKAAVEMFGKDFQGS 120
379 DFFKQGVVNMNRKTQPMIHYLDKETGPKGDAATVSYEDPTAKAAVEMFGKDFQGS 438
121 KLVSLARKKPPMNSMRGGLPPREGGMPPLRLGPGGPGGPGMGRGGRGDRGFP 180
439 KLVSLARKKPPMNSMRGGLPPREGGMPPLRLGPGGPGGPGMGRGGRGDRGFP 498
181 PRGPRGSRGNPSGGVNVQHRAGDMQCPNPGCGNQNFMARTECNQCAKPKREGFLPPFP 240
499 PRGPRGSRGNPSGGVNVQHRAGDMQCPNPGCGNQNFMARTECNQCAKPKREGFLPPFP 558
241 PGDGRGPGGPGMGRGGRGGLMDRGPGCMFRGGRGGRGGRGGRGGRGGRG 300
559 PGDGRGPGGPGMGRGGRGGLMDRGPGCMFRGGRGGRGGRGGRGGRGGRG 618

QY 301 GGPPLMEQMGRRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRG 338
DB 619 GGPPLMEQMGRRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRG 656

RESULT 3

US-10-755-889-650
Sequence 650, Application US/10755889
Publication No. US20040171823A1
GENERAL INFORMATION:
APPLICANT: Bristol-Myers Squibb Company
TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES ASSOCIATED WITH THE NF-KB
FILE REFERENCE: D0284 NP
CURRENT FILING DATE: US/10/755, 889
PRIOR FILING DATE: 2004-01-13
PRIOR APPLICATION NUMBER: U.S. 60/440,068
PRIOR FILING DATE: 2003-01-14
PRIOR APPLICATION NUMBER: U.S. 60/469,757
SOFTWARE: PatentIn version 3.2
NUMBER OF SEQ ID NOS: 823
SEQ ID NO 650
LENGTH: 525
TYPE: PRT
ORGANISM: Homo sapiens
US-10-755-889-650

Query Match 45.7%; Score 878; DB 16; Length 525;
Best Local Similarity 52.9%; Pred. No. 2, 9e-54;
Matches 182; Conservative 36; Mismatches 54; Indels 72; Gaps 13;

1 GGRGMSAGERGPFKPGPMDEGPDLDLGPVDDPDESDNSAIYVGLNDSVTLLDILA 60
248 GGRGMSAGERGPFKPGPMDEGPDLDLGPVDDPDESDNSAIYVGLNDSVTLLDILA 301
61 DFFKQGVVNMNRKTQPMIHYLDKETGPKGDAATVSYEDPTAKAAVEMFGKDFQGS 120
302 DFFKQGVVNMNRKTQPMIHYLDKETGPKGDAATVSYEDPTAKAAVEMFGKDFQGS 361
121 KLVSLARKKPPMNSMRGGLPPREGGMPPLRLGPGGPGGPGMGRGGRGDRGFP 180
362 KLVSLARKKPPMNSMRGGLPPREGGMPPLRLGPGGPGGPGMGRGGRGDRGFP 410
181 PRGPRGSRGNPSGGVNVQHRAGDMQCPNPGCGNQNFMARTECNQCAKPKREGFLPPFP 240
411 PRGPRGSRGNPSGGVNVQHRAGDMQCPNPGCGNQNFMARTECNQCAKPKREGFLPPFP 453
241 PGDGRGPGGPGMGRGGRGGLMDRGPGCMFRGGRGGRGGRGGRGGRGGRG 295
454 PGDGRGPGGPGMGRGGRGGLMDRGPGCMFRGGRGGRGGRGGRGGRGGRG 505
296 RRGPGGPGGPGMGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRG 338
506 RRGPGGPGGPGMGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRG 525

RESULT 4

US-10-408-765A-116
Sequence 116, Application US/10408765A
Publication No. US20040101874A1
GENERAL INFORMATION:
APPLICANT: Ghosh, Soumitra S.
APPLICANT: Fahy, Eoin D.
APPLICANT: Zhang, Bing
APPLICANT: Gibson, Bradford W.
APPLICANT: Taylor, Steven W.
APPLICANT: Glenn, Gary W.
APPLICANT: Wainock, Dale E.
TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION
TITLE OF INVENTION: IDENTIFIED IN THE MITOCHONDRIAL PROTEOME
FILE REFERENCE: 660088.465

CURRENT APPLICATION NUMBER: US/10/408,765A
CURRENT FILING DATE: 2003-04-04
NUMBER OF SEQ ID NOS: 3077
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 116
LENGTH: 260
TYPE: PRT
ORGANISM: Homo sapiens
US-10-408-765A-116

Query Match 41.6%; Score 799.5; DB 16; Length 260;
Best Local Similarity 51.4%; Pred. No. 5,3e-49;
Matches 167; Conservative 34; Mismatches 53; Indels 71; Gaps 12;

QY 20 GPMDEGPDLDLGPVDEDEDSNLSAIYVGLINDSVTLDDIADPFKQGVVKNKRTGQPM 79
DB 1 GPRDQGSRRHD-----SEQDNDNNITIFVQGLGEVITLESVADYFKQIGIKTKWKTGQPM 55
QY 80 IHTYLDKENTGKPKGDATVSYEDPTAKAAYEWPDGKDFQSGSKLVSLARKKPPNMSRGG 139
DB 56 IHTYLDKENTGKPKGDATVSYEDPTAKAAYEWPDGKDFQSGSKLVSLARKKPPNMSRGG 113
QY 140 LPPREGGMPPLRGFGGPGGPGGPGGPGGPGGPGGPGGPGGPGGPGGPGGPGGPGGPGG 199
DB 114 GNGRGGRG-----RGGMGBGGYGGG-----GSGGGGGRGFPSSG-----GGGGGQ 155
QY 200 RAGDMQCPNPGCGNPFAMTECNQCAPPEGFLPPPPPGGDRGGRGPGGPGGPGGPGGPGG 258
DB 156 RAGDMQCPNPGCGNPFAMTECNQCAPPEGFLPPPPPGGDRGGRGPGGPGGPGGPGGPGG 202
QY 259 GLMDRGPGGMPRG-----GSGGDRGPGGPGGPGGPGGPGGPGGPGGPGGPGGPGGPGG 314
DB 203 GDRRGGRGGYDGGYRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGR 240
QY 315 RGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGR 338
DB 241 -----GPGKMDSRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGR 260

RESULT 5

US-09-925-301-1511
Sequence 1511, Application US/09925301
Patent No. US20020052308A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
FILE REFERENCE: PA106
CURRENT APPLICATION NUMBER: US/09/925,301
PRIOR APPLICATION NUMBER: PCT/US00/05682
PRIOR FILING DATE: 2001-08-10
PRIOR FILING DATE: 2000-03-08
PRIOR APPLICATION NUMBER: 60/124,270
PRIOR FILING DATE: 1999-03-12
NUMBER OF SEQ ID NOS: 1694
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1511
LENGTH: 156
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: SITE
LOCATION: (11)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
LOCATION: (104)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: SITE
LOCATION: (143)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-301-1511

Query Match 32.4%; Score 623; DB 9; Length 156;
Best Local Similarity 73.2%; Pred. No. 1e-36;

Matches 115; Conservative 1; Mismatches 7; Indels 34; Gaps 4;

QY 160 GSGFGPMGRMGSGDRGPGPPPRGPRGSRGNPSGGNVQHRADWQCPNP----- 209
DB 18 GSPG--LQERGTR--DRGFPPRGPRGSRGNPSGGNVQHRADWQCPNPSIGDFCCDVI 73
QY 210 ---GCCGNQNPAMTECNQCAPPEGFLPPPPPGGDRGGRGGRGGRGGRGGRGGRGGR 266
DB 74 VCRGCGNQPAMTECNQ-----GDRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGR 116
QY 267 GGMFRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGR 303
DB 117 GGMFRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGR 153

RESULT 6

US-09-864-761-42397
Sequence 42397, Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE REFERENCE: Aeomica-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
PRIOR FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263.6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
SEQ ID NO 42397
LENGTH: 83
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AC000026.3

OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.2
 OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.9
 OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.3
 OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 2
 OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.5
 OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 3.5
 OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 5.8
 OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.7
 US-09-864-761-42397

Query Match 24.7%; Score 475; DB 9; Length 83;
 Best Local Similarity 100.0%; Pred. No. 1.7e-26;
 Matches 83; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 243 GDRGRGPGMGRGGLMDRGGMFRGGRGDRGFRGGRGMDRGGFRGGRGPG 302
 DB 1 GDRGRGPGMGRGGLMDRGGMFRGGRGDRGFRGGRGMDRGGFRGGRGPG 60
 QY 303 PPGPLMEQMGRRGRGGRGPGMD 325
 DB 61 PPGPLMEQMGRRGRGGRGPGMD 83

RESULT 7

US-10-437-963-191926
 Sequence 191926, Application US/10437963
 Publication No. US20040123343A1

GENERAL INFORMATION:
 APPLICANT: La Rosa, Thomas J.
 APPLICANT: Kovalic, David K.
 APPLICANT: Zhou, Yihua
 APPLICANT: Cao, Yongwei
 APPLICANT: Wu, Wei
 APPLICANT: Boukharov, Andrey A.
 APPLICANT: Barbazuk, Brad
 TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
 TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
 FILE REFERENCE: 38-21(53221)B
 CURRENT APPLICATION NUMBER: US/10/437,963
 CURRENT FILING DATE: 2003-05-14
 NUMBER OF SEQ ID NOS: 204966
 SEQ ID NO 191926
 LENGTH: 467
 TYPE: PRT
 ORGANISM: Oryza sativa
 FEATURE:
 OTHER INFORMATION: Clone ID: PAT_MRT4530_881C.1.pap
 US-10-437-963-191926

Query Match 22.0%; Score 423; DB 16; Length 467;
 Best Local Similarity 32.4%; Pred. No. 4.9e-22;
 Matches 113; Conservative 29; Mismatches 77; Indels 130; Gaps 12;

QY 31 GPPVDPEDSDNSAIYVQGLNDSVTLDDLADFFKQCGVVKKRTGQPMIHYLDKETGK 90
 DB 95 GPP-----NSVYVCNLPFGDTERMLADYFETTLKDKKRTGPKMITYRDKTYNE 146
 QY 91 PKGDATVSYEDPPTAKAAYVWFDGKDFQSGSLKLYSLARKK-----PPMNSMRGGLPP 142
 DB 147 PKGDATVSYEDPPTAKAAYVWFDGKDFQSGSLKLYSLARKK-----PPMNSMRGGLPP 142
 QY 143 RGRGMPPLRGGGCGGCGGPGMGRGGR-----GGDRGFRGGRGGRGGRGPGSGGCV 197
 DB 203 -----LGQDELNDNGARG-----RGHGDGPKA 226
 QY 198 QHRAGDMQCPNPGCGNPFAMRTECNQCKAPKEGFLPPFPFPFGGDRGGRGGRGGR 257
 DB 227 WQDDGMLCINTSGCNVNFAPRGVNCNRGAARPG----- 261
 QY 258 GGLMDRGGPGMGRGGRGDRGFRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGR 308
 DB 262 ---VSGSGAGG---GGRG-----RGRGSD-DAKGSRAAAVAGPGLFGRPDWGCPMC 307

QY 309 EGM-----GRRGRGGRGPGK-MDKGRHQRDRR 336
 DB 308 GNINAKMKKNCINTTKRHNHGGVGRGGGYKELDEBELEEVKRR 356

RESULT 8

US-10-425-114-69744
 Sequence 69744, Application US/10425114
 Publication No. US20040034888A1

GENERAL INFORMATION:
 APPLICANT: Liu, Jindong
 APPLICANT: Zhou, Yihua
 APPLICANT: Kovalic, David K.
 APPLICANT: Screen, Steven E.
 APPLICANT: Tabaska, Jack E.
 APPLICANT: Cao, Yongwei
 TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
 TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
 FILE REFERENCE: 38-21(53313)B
 CURRENT APPLICATION NUMBER: US/10/425,114
 CURRENT FILING DATE: 2003-04-28
 NUMBER OF SEQ ID NOS: 73128
 SEQ ID NO 69744
 LENGTH: 395
 TYPE: PRT
 ORGANISM: Zea mays
 FEATURE:
 OTHER INFORMATION: Clone ID: UC-ZMFLB73242A07_FLI.pap
 US-10-425-114-69744

Query Match 21.9%; Score 420; DB 15; Length 395;
 Best Local Similarity 32.4%; Pred. No. 6.7e-22;
 Matches 113; Conservative 31; Mismatches 95; Indels 110; Gaps 11;

QY 20 GPMDEGPDLDLPPVDPEDSDNSAIYVQGLNDSVTLDDLADFFKQCGVVKKRTGQPM 79
 DB 12 GPMDEGPDLDLPPVDPEDSDNSAIYVQGLNDSVTLDDLADFFKQCGVVKKRTGQPM 79
 QY 80 IHIYDKETGKRGDATTYEDPPTAKAAYVWFDGDFQSGSLKLYSLARKKPP----- 132
 DB 58 IHIYDKETGKRGDATTYEDPPTAKAAYVWFDGDFQSGSLKLYSLARKKPP----- 132
 QY 133 -MNSMRGGLPPRGRGMPPLRGGGCGGPGMGRGGRGGRGGRGGRGGRGGRGGRGGR 191
 DB 118 -MNSMRGGLPPRGRGMPPLRGGGCGGPGMGRGGRGGRGGRGGRGGRGGRGGRGGR 191
 QY 192 SGGGVQHRAGDMQCPNPGCGNPFAMRTECNQCKAPKEGFLPPFPFPFGGDRGGRGPG 251
 DB 143 SGGGVQHRAGDMQCPNPGCGNPFAMRTECNQCKAPKEGFLPPFPFPFGGDRGGRGPG 251
 QY 252 GMRGGRGGLMDRGGPGMGRG--GRGDRGFRGGRGGRGGRGGRGGRGGRGGRGGR 308
 DB 189 GMRGGRGGLMDRGGPGMGRG--GRGDRGFRGGRGGRGGRGGRGGRGGRGGRGGR 308
 QY 309 EGM-----GRRGRGGRGPGK-MDKGRHQRDRR 336
 DB 234 GNINAKMKKNCINTSKRPTNHEGVRGGRGGYKELDEBELEEVKRR 262

RESULT 9

US-10-437-963-164510
 Sequence 164510, Application US/10437963
 Publication No. US20040123343A1

GENERAL INFORMATION:
 APPLICANT: La Rosa, Thomas J.
 APPLICANT: Kovalic, David K.
 APPLICANT: Zhou, Yihua
 APPLICANT: Cao, Yongwei
 APPLICANT: Wu, Wei
 APPLICANT: Boukharov, Andrey A.
 APPLICANT: Barbazuk, Brad
 APPLICANT: Li, Ping


```

/ TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated with
/ FILE REFERENCE: 38-21(53221)B
/ CURRENT APPLICATION NUMBER: US/10/437,963
/ NUMBER OF SEQ ID NOS: 204966
/ SEQ ID NO 164510
/ LENGTH: 578
/ TYPE: PRT
/ ORGANISM: Oryza sativa
/ FEATURE:
/ OTHER INFORMATION: Clone ID: PAT_MRT4530_63401C.1.pep
US-10-437-963-164510

Query Match          16.6%; Score 319.5; DB 16; Length 578;
Best Local Similarity 3.5%; Pred. No. 1.3e-14;
Matches 90; Conservative 9; Mismatches 74; Indels 67; Gaps 8;

QY 145 GRGMPPLRGGPGGPGGPGMGRMGRGG-----DRGPPPRGPRGRGNBPGGGNVQ 198
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 64 GGGVGGGCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 122

QY 199 HRAADMCCPNP-----GCCGNON 215
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 123 GRDGDWVCPDPRILHGACVACPTKFLFHRGINIRLQFHFSSNIPNIFVYHSSCGNVN 182

QY 216 FAARTENOCARKEPEFLPPPPPPGDDRGPGGMRGGRGLMDRGPGGMRGGRG 275
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 183 FAARTENCKGAPSPAG-----GGGGGGGGGGGKSGGGGGGYNRGG-GDFSSSGGGG 232

QY 276 G-DRGG---FRGGRGMDRGFGGGGRRGGPGGPPPLMEQMGRRGGRGGGGKXDKGHRQ 331
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 233 GYNRGGDVNSGGG---GGTGGGGGKGGYNRGGGDDRGDFDHRGGGGYGGRDQGNQR 289

```

```

RESULT 10
US-10-424-599-187688
/ Sequence 187688, Application US/10424599
/ Publication NO. US20040031072A1
/ GENERAL INFORMATION:
/ APPLICANT: La Rosa Thomas J
/ APPLICANT: Kovalic David K
/ APPLICANT: Zhou Yihua
/ APPLICANT: Cao Yongwei
/ TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated with
/ FILE REFERENCE: 38-21(53223)B
/ CURRENT APPLICATION NUMBER: US/10/424,599
/ CURRENT FILING DATE: 2003-04-28
/ NUMBER OF SEQ ID NOS: 285684
/ SEQ ID NO 187688
/ LENGTH: 197
/ TYPE: PRT
/ ORGANISM: Glycine max
/ FEATURE:
/ NAME/KEY: unsure
/ LOCATION: (1)..(197)
/ OTHER INFORMATION: unsure at all Xaa locations
/ FEATURE:
/ OTHER INFORMATION: Clone ID: PAT_MRT3847_140496C.1.pep
US-10-424-599-187688

```

```

Query Match          15.8%; Score 303; DB 15; Length 197;
Best Local Similarity 3.4%; Pred. No. 6.6e-14;
Matches 79; Conservative 21; Mismatches 70; Indels 60; Gaps 7;

QY 42 NSAIYQGLSDSTLDD--LADFFKQCGVYKMKRTGQPMIHIYLDKRTGPKRGDATTVEY 99
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 12 NSGVYV--CNLPYGTDDIMLAIFYGTIGLTKDKRTGRPTWLRKKTETNPKGDATTIVY 69

QY 100 EDPPTAKAAVEMWDGDFQSGSKLVSLARKK-----PMMNSMRGGLPPREGRG 147
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 70 EDPHAAVAALVEMWNNNDFTGNTIGTIVLAEISKNDDEQAYNAAVEPVADVGL-----EE 124

```

```

QY 148 MPEPLRGPGPGPGPGPMGRMGRGGDRGGFPPEPRGRGRGNBPGGGNVQHRAGDWCP 207
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 125 TTKDVGGSG-----RGXXONDS--SGKXWQDGDWICL 156

QY 208 NPGGNGONFAMRTECNOCARKEPEFLPPPPPPGDDRGPGGMRGGR 257
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 157 NTGCSNVNFAFRACNKRXTAR-----SAGASGISGAGRGKR 195

```

```

RESULT 11
US-10-058-124-21
/ Sequence 21, Application US/10058124
/ Publication NO. US20030119058A1
/ GENERAL INFORMATION:
/ APPLICANT: Qvist, Per
/ APPLICANT: Bonde, Martin
/ TITLE OF INVENTION: A Method for Assaying Collagen Fragments
/ in Body Fluids, A Test Kit and Means for Carrying Out the
/ Method and Use of the Method to Diagnose the Presence of
/ Disorders Associated with the Metabolism of

NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESSES:
ADDRESSER: Darby & Darby PC
STREET: 805 Third Avenue
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10022

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.25

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/058,124
FILING DATE: 29-Jan-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/570,573
FILING DATE: 2002-MAY-12
APPLICATION NUMBER: 08/187,319
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Gogoris, Adda C
REGISTRATION NUMBER: 29,714
REFERENCE/DOCKET NUMBER: 4305/08701
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-527-7700
TELEFAX: 212-753-6237
TELEX: 236687

INFORMATION FOR SEQ ID NO: 21:
SEQUENCE CHARACTERISTICS:
LENGTH: 1078 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
IMMEDIATE SOURCE:
CLONE: COLLAGEN ALPHA 1 (III)
SEQUENCE DESCRIPTION: SEQ ID NO: 21:
US-10-058-124-21

```

```

Query Match          15.8%; Score 303; DB 14; Length 1078;
Best Local Similarity 30.3%; Pred. No. 3.7e-13;
Matches 115; Conservative 15; Mismatches 126; Indels 124; Gaps 16;

QY 1 GRRGGSAGRGGRGGRGPMDEGPDLDGPPVDPDESDNSAIYVQGLSDSTLDDLA 60
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 291 GAKGSGRGRGGRGEGIGIPGVGAKGDEKDGSPDGGANGLGAGGERALDS----- 343

QY 61 DFKQCGVYKMKRTGQPMIHIYLDKRTGPKRGDATTVEDEPTAKAAVEMWDGDFQGS 120

```

Db 344 -----RCPAPNPIPEKPGAG----- 361
 QY 121 KUKVSLARKKPPNNMNRG--GLPREGKMPPLRGCGPGGPG--GPMKRMG--G 171
 Db 362 -----RGAPGACRGGAAGBPGRDGVPGCGPMRGPGSGPGSDGKPGPGSGEGS 414
 QY 172 RGGDRGPPRRGSRG-----NPSGGGVCHRADWMCNPGCNQNPAMTECNQCK 226
 Db 415 RPBPFG--PSGRGPGVWGFPGPKNDGAPKKNBERGGPG--GPPBQG----- 459
 QY 227 APKEGFLPPFPFPP-----PGDGRGPGGPMRG-----GRGIMDR----- 263
 Db 460 PPKKNBYEPGPGPGPTGPGDGDGDTGPRPGQLGGLPRTGSGPGENEKRGEPKGEAG 519
 QY 264 --GGPGMFRGGRGDRG-----GPRGGRMDRGFGGRRGPGGPG--PLME 309
 Db 520 APGAPGKGDAAGAPRGPPGLAGAPGLRGAG--PGEBGKGGAAGPPGPGAGTPTGLQ 578
 QY 310 QMGRRGGRGPG--KMDKGE 328
 Db 579 GMPERGGLGSGPGPKDCKB 598

RESULT 12

US-10-402-089-4
 ; Sequence 4, Application US/10402089
 ; Publication No. US20040005663A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Bell, Marcum P.
 ; APPLICANT: Neff, Thomas B.
 ; APPLICANT: Polarek, James W.
 ; APPLICANT: Sealey, Todd W.
 ; TITLE OF INVENTION: PORCINE COLLAGENS AND GELATINS
 ; FILE REFERENCE: PP0402.3 CON
 ; CURRENT APPLICATION NUMBER: US/10/402,089
 ; CURRENT FILING DATE: 2003-03-26
 ; PRIOR APPLICATION NUMBER: US 09/709,700
 ; PRIOR FILING DATE: 2000-11-10
 ; NUMBER OF SEQ ID NOS: 72
 ; SOFTWARE: PatentIn version 3.2
 ; SEQ ID NO 4
 ; LENGTH: 1466
 ; TYPE: PRT
 ; ORGANISM: Bos Taurus
 ; US-10-402-089-4

Query Match 15.8%; Score 303; DB 15; Length 1466;
 Best Local Similarity 31.2%; Pred. No. 5.1e-13;
 Matches 130; Conservative 22; Mismatches 157; Indels 108; Gaps 21;

QY 1 GGRGKMSAGBRGCFNKPGGPMDEGPDLDLGPVPDDESDNSAIYVQGLNDSVTLDDL 60
 Db 348 GAKGEVGPAGSPGSSGAPGQGRGEPGQSHAGAPGPBPSPGNSGPGKGEKMGAPGAP 407
 QY 61 DFFKQGVVKNKRTGQPMHIYLDK-----ETGK--PKGDATVSYE-----DPTAKAA 108
 Db 408 -----GLIGARGPPEPTNGVPGQGAAGEPKKNAKADPPGRRGERGAAGBGLAGPK 461
 QY 109 VEWFDGKD-----FOGSKLVSLARKKPPNNMNRG----- 138
 Db 462 GE--DGKDSGPERGANGLPGAAGERVVPGRBPAGANGLPGEKGPFGDGGGPPAPRG 519
 QY 139 --GLPREGKMPPLRGCGPGGPG--GPMKRMG-----GRGDRG-----G 178
 Db 520 VAGPGRDGLPGGGLRGLRGSPGSDGKPGPPGQGETGRBPSPGSPGPPGQPGVWG 579
 QY 179 FPPRGPRGSGNPS-----GGGVNHRAGDMQCPNPGCGNQNFAMRTCCNCKA--- 227
 Db 580 FP--GPKNDGAPKKNBERGGPGPGPG--KNGETGPGPGPTGPGSDGKDTG 633
 QY 228 -PKREGF--LPPFPFPG-----GDRGGRGPGMRGGRGLMDRGCGPMFRGGRG 276

Db 634 PPGQGLGGLPTGSPGPNKRGKPGKGAAGAPGIPGKGG---DSGAPGRGPGGAG 690
 QY 277 DRGFRGRGMRGRGFRGGRGPGGPG--PLMEOWGRRGRGPG--KMDKGE 328
 Db 691 PPGP--RGAG--PPGEGKGGAAPPGPPGSPGACTPGLQMPGRGGPGGPGKDKB 745

RESULT 13

US-10-402-089-6
 ; Sequence 6, Application US/10402089
 ; Publication No. US20040005663A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Bell, Marcum P.
 ; APPLICANT: Neff, Thomas B.
 ; APPLICANT: Polarek, James W.
 ; APPLICANT: Sealey, Todd W.
 ; TITLE OF INVENTION: PORCINE COLLAGENS AND GELATINS
 ; FILE REFERENCE: PP0402.3 CON
 ; CURRENT APPLICATION NUMBER: US/10/402,089
 ; CURRENT FILING DATE: 2003-03-26
 ; PRIOR APPLICATION NUMBER: US 09/709,700
 ; PRIOR FILING DATE: 2000-11-10
 ; NUMBER OF SEQ ID NOS: 72
 ; SOFTWARE: PatentIn version 3.2
 ; SEQ ID NO 6
 ; LENGTH: 1466
 ; TYPE: PRT
 ; ORGANISM: Sus scrofa
 ; US-10-402-089-6

Query Match 15.8%; Score 303; DB 15; Length 1466;
 Best Local Similarity 31.2%; Pred. No. 5.1e-13;
 Matches 130; Conservative 22; Mismatches 157; Indels 108; Gaps 21;

QY 1 GGRGKMSAGBRGCFNKPGGPMDEGPDLDLGPVPDDESDNSAIYVQGLNDSVTLDDL 60
 Db 348 GAKGEVGPAGSPGSSGAPGQGRGEPGQSHAGAPGPBPSPGNSGPGKGEKMGAPGAP 407
 QY 61 DFFKQGVVKNKRTGQPMHIYLDK-----ETGK--PKGDATVSYE-----DPTAKAA 108
 Db 408 -----GLIGARGPPEPTNGVPGQGAAGEPKKNAKADPPGRRGERGAAGBGLAGPK 461
 QY 109 VEWFDGKD-----FOGSKLVSLARKKPPNNMNRG----- 138
 Db 462 GE--DGKDSGPERGANGLPGAAGERVVPGRBPAGANGLPGEKGPFGDGGGPPAPRG 519
 QY 139 --GLPREGKMPPLRGCGPGGPG--GPMKRMG-----GRGDRG-----G 178
 Db 520 VAGPGRDGLPGGGLRGLRGSPGSDGKPGPPGQGETGRBPSPGSPGPPGQPGVWG 579
 QY 179 FPPRGPRGSGNPS-----GGGVNHRAGDMQCPNPGCGNQNFAMRTCCNCKA--- 227
 Db 580 FP--GPKNDGAPKKNBERGGPGPGPG--KNGETGPGPGPTGPGSDGKDTG 633
 QY 277 DRGFRGRGMRGRGFRGGRGPGGPG--PLMEOWGRRGRGPG--KMDKGE 328
 Db 691 PPGP--RGAG--PPGEGKGGAAPPGPPGSPGACTPGLQMPGRGGPGGPGKDKB 745

RESULT 14

US-10-402-072A-4
 ; Sequence 4, Application US/10402072A
 ; Publication No. US20040018592A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Bell, Marcum P.
 ; APPLICANT: Neff, Thomas B.
 ; APPLICANT: Polarek, James W.
 ; APPLICANT: Sealey, Todd W.
 ; TITLE OF INVENTION: BOVINE COLLAGENS AND GELATINS

FILE REFERENCE: PP0402.2 CON
 CURRENT APPLICATION NUMBER: US/10/402,072A
 CURRENT FILING DATE: 2003-03-26
 PRIOR APPLICATION NUMBER: US 09/709,700
 PRIOR FILING DATE: 2000-11-10
 NUMBER OF SEQ ID NOS: 72
 SOFTWARE: PatentIn version 3.2
 SEQ ID NO 4
 LENGTH: 1466
 TYPE: PRT
 ORGANISM: Bos Taurus
 US-10-402-072A-4

Query Match 15.8%; Score 303; DB 15; Length 1466;
 Best Local Similarity 31.2%; Pred. No. 5.1e-13;
 Matches 130; Conservative 22; Mismatches 157; Indels 108; Gaps 21;

QY 1 GGRGMSAGRGGRGFFNKGPMDEGPDLDGPPVDPEDSDNSAIYVQGLNDSTVTLDDLA 60
 DB 348 GAKGEVGPAGSPSSSGAPGQGRGEPFGHAGAPPPPPSGNSPGGKGEKGPAGIPGAP 407
 QY 61 DFFKQCGVVMKRRKTGQPMIHYLDK-----ETGK--PKGDATVSYE-----DPTAKAA 108
 DB 408 -----GLIGARPPGPGTGVPGQGAAGEPGKNAKDDPPRGREGAGSGFTIAGPK 461
 QY 109 VEMFDGKD-----FQSKLVSLARKKPPMNSMRG----- 138
 DB 462 GE--DGKDGSPGBPGANGLPGAAGRGVPPFRGPAGANGLPGEKGPDRGPGPAPRG 519
 QY 139 --GLPPREGKMPPLRGPGGPGPG-----GPMGRMG-----GRGDDG-----G 178
 DB 520 VAGEPGDGLPGGTGLRGTSPGPGSDGKPPGSGQGETGRPPGSPGPGQPGVWG 579
 QY 179 FPPRGPRGSRGNPS-----GGNVQHRAGDMQCPNPGCGNQNFAMRTECNQCKA--- 227
 DB 580 FP--GPKNDGAPKNGERGGPGGPPGQPGAG---KNGETGPGGPPGPTGPGSDKDDTG 633
 QY 228 -PKREGF--LPPPPPPPG-----GDRGRGPGGMRGGRGGLMDRGPGGMFRGGRG 276
 DB 634 PPGPQGLQGLPGTSGPPGKGPBPGPKGEAGAPGIPGKG---DSGAPGBRGPPGAGG 690
 QY 277 DRGGRGGRGMDRGGRGGRGPGPGPG---PLMEQMGRRGGRGPGG-KMDKGE 328
 DB 691 PPGP-RGGAG-PPGPGGKGAAGPPPPGSACTGTGLQGMGERGGPGGPPKDKGE 745

RESULT 15

US-10-402-072A-6
 Sequence 6, Application US/10402072A
 Publication No. US20040018592A1
 GENERAL INFORMATION:
 APPLICANT: Bell, Marcum P.
 APPLICANT: Naff, Thomas B.
 APPLICANT: Polarek, James W.
 APPLICANT: Sealey, Todd W.
 TITLE OF INVENTION: BOVINE COLLAGENS AND GELATINS
 FILE REFERENCE: PP0402.2 CON
 CURRENT APPLICATION NUMBER: US/10/402,072A
 CURRENT FILING DATE: 2003-03-26
 PRIOR APPLICATION NUMBER: US 09/709,700
 PRIOR FILING DATE: 2000-11-10
 NUMBER OF SEQ ID NOS: 72
 SOFTWARE: PatentIn version 3.2
 SEQ ID NO 6
 LENGTH: 1466
 TYPE: PRT
 ORGANISM: Sus scrofa
 US-10-402-072A-6

Query Match 15.8%; Score 303; DB 15; Length 1466;
 Best Local Similarity 31.2%; Pred. No. 5.1e-13;
 Matches 130; Conservative 22; Mismatches 157; Indels 108; Gaps 21;

QY 1 GGRGMSAGRGGRGFFNKGPMDEGPDLDGPPVDPEDSDNSAIYVQGLNDSTVTLDDLA 60
 DB 348 GAKGEVGPAGSPSSSGAPGQGRGEPFGHAGAPPPPPSGNSPGGKGEKGPAGIPGAP 407
 QY 61 DFFKQCGVVMKRRKTGQPMIHYLDK-----ETGK--PKGDATVSYE-----DPTAKAA 108
 DB 408 -----GLIGARPPGPGTGVPGQGAAGEPGKNAKDDPPRGREGAGSGFTIAGPK 461
 QY 109 VEMFDGKD-----FQSKLVSLARKKPPMNSMRG----- 138
 DB 462 GE--DGKDGSPGBPGANGLPGAAGRGVPPFRGPAGANGLPGEKGPDRGPGPAPRG 519
 QY 139 --GLPPREGKMPPLRGPGGPGPG-----GPMGRMG-----GRGDDG-----G 178
 DB 520 VAGEPGDGLPGGTGLRGTSPGPGSDGKPPGSGQGETGRPPGSPGPGQPGVWG 579
 QY 179 FPPRGPRGSRGNPS-----GGNVQHRAGDMQCPNPGCGNQNFAMRTECNQCKA--- 227
 DB 580 FP--GPKNDGAPKNGERGGPGGPPGQPGAG---KNGETGPGGPPGPTGPGSDKDDTG 633
 QY 228 -PKREGF--LPPPPPPPG-----GDRGRGPGGMRGGRGGLMDRGPGGMFRGGRG 276
 DB 634 PPGPQGLQGLPGTSGPPGKGPBPGPKGEAGAPGIPGKG---DSGAPGBRGPPGAGG 690
 QY 277 DRGGRGGRGMDRGGRGGRGPGPGPG---PLMEQMGRRGGRGPGG-KMDKGE 328
 DB 691 PPGP-RGGAG-PPGPGGKGAAGPPPPGSACTGTGTGLQGMGERGGPGGPPKDKGE 745

Search completed: February 18, 2005, 15:43:17
 Job time : 143 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 18, 2005, 15:22:27 / Search time 40 Seconds

(without alignments)
813.032 Million cell updates/sec

Title: US-10-791-017a-2_COPY_319_656

Perfect score: 1922

Sequence: 1 GGRGMSAGSRGPFKPGG.....GGPGKMDKGRHRRDRPV 338

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1922	100.0	656	1 A49358	RNA-binding protei
2	1888.5	98.3	655	1 A55726	RNA-binding protei
3	878	45.7	526	1 S33799	RNA-binding protei
4	871	45.3	528	2 G02127	fus-like protein -
5	750	39.0	589	2 S71954	RNA/sedna-binding
6	623.5	32.4	404	2 S54729	RNA-binding protei
7	420.5	21.9	545	2 T15667	hypothetical prote
8	370.5	19.3	211	2 C96539	hypothetical prote
9	302	15.7	886	2 I50694	collagen alpha 1(I
10	301	15.7	1466	1 CGHUTL	collagen alpha 1(I
11	299.5	15.6	1049	1 CGBO78	collagen alpha 1(I
12	298	15.5	310	2 I50696	collagen alpha 1(I
13	293	15.2	1464	2 S59856	collagen alpha 1(I
14	291	15.1	276	2 T33925	hypothetical prote
15	283.5	14.8	1042	1 CGCH1S	collagen alpha 1(I
16	278.5	14.5	326	2 A41732	heterogeneous ribo
17	277.5	14.4	1147	1 MMXIB	myosin heavy chain
18	275.5	14.3	1486	1 CGH2V	collagen alpha 2(V
19	271.5	14.1	386	1 S22315	anRNP-associated p
20	270.5	14.1	940	2 JEO291	FB19 protein - hum
21	270	14.0	1419	2 A41182	collagen alpha 1(I
22	270	14.0	1487	2 B41182	collagen alpha 1(I
23	270	14.0	1492	2 A40333	collagen alpha 1(I
24	270	14.0	1497	2 I49607	procollagen type V
25	269.5	14.0	342	2 S14432	heterogeneous ribo
26	269	14.0	365	2 A26459	helix-desfibrillin
27	269	14.0	1486	1 B40333	collagen alpha 1(I
28	268.5	14.0	1758	2 T29350	hypothetical prote
29	265.5	13.8	1453	2 S21626	collagen alpha 1(I

30	265	13.8	352	2 T24279	hypothetical prote
31	265	13.8	1603	2 S23810	collagen alpha 1(X
32	264	13.7	435	2 T15143	hypothetical prote
33	263.5	13.7	1464	1 CGH01S	collagen alpha 1(I
34	263.5	13.7	1880	2 T18531	tracrin - medicina
35	262	13.6	414	2 JN0866	nucleolar protein
36	262	13.6	3176	2 CGH03A	collagen alpha 3(V
37	261.5	13.6	671	1 CGRT1S	collagen alpha 1(I
38	261.5	13.6	779	1 CGB01S	collagen alpha 1(I
39	261.5	13.6	1418	2 T45467	collagen alpha 1(I
40	261.5	13.6	1806	2 CGH01E	collagen alpha 1(X
41	261	13.6	2944	2 A54849	collagen alpha 1(V
42	259.5	13.5	309	2 T19389	collagen alpha 1(V
43	259	13.5	346	1 S35500	hypothetical prote
44	259	13.5	385	2 S35500	heterogeneous ribo
45	258.5	13.4	1487	1 CGH06C	ribonucleoprotein
					collagen alpha 1(I

ALIGNMENTS

RESULT 1
A49358
RNA-binding protein EMS - human
C:Species: Homo sapiens (man)
C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 09-Jul-2004
C/Accession: A49358; S28257
R;Plogastel, B.; Zucman, J.; Peter, M.; Thomas, G.; Delattre, O.
Genomics 18, 609-615, 1993
A>Title: Genomic structure of the EMS gene and its relationship to EMSR1, a site of tum
A/Reference number: A49358; MIMD:9410360; PMID:8307570
A/Accession: A49358
A/Status: nucleic acid sequence not shown; translation not shown
A/Molecule type: DNA
A/Residues: 1-656 <RES>
A/Cross-references: UNIPROT:O01844; EMBL:X72990; NID:9485838; PID:CAA51489.1; PID:9825.
R;Delattre, O.; Zucman, J.; Plogastel, B.; Desmarte, C.; Melot, T.; Peter, M.; Kovar, H
Nature 359, 162-165, 1992
A>Title: Gene fusion with an ETS DNA-binding domain caused by chromosome translocation
A/Reference number: S28257; MIMD:92396239; PMID:1522903
A/Accession: S28257
A/Molecule type: mRNA
A/Residues: 1-656
A/Cross-references: EMBL:X66899; NID:9547565; PID:CAA47350.1; PID:G31280
C/Genetics:
A/Genes: GDB:EMSR1
A/Cross-references: GDB:135984; OMIM:133450
A/Map position: 22q12.1-22q12.1
A/Intons: 5/1; 17/2; 34/3; 76/1; 138/2; 194/2; 265/1; 325/2; 338/1; 349/1; 388/3; 432/
A/Note: EMSR1 region is exons 7-10 of this gene, called EMS in reference A49358; this r
C/Superfamily: RNA-binding protein, EMS type; ribonucleoprotein repeat homology
C/Keywords: carcinogenesis; nucleus; RNA binding; tandem repeat
F;362-437/Domain: ribonucleoprotein repeat homology <RRM>

Query Match
Best Local Similarity 100.0%; Score 1922; DB 1; Length 656;
Matches 338; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	GGRGMSAGSRGPFKPGGPMDEGPDLDGPPVDDDESDNSATVVGGLNDSVTLDLLA	60
DB	319	GGRGMSAGSRGPFKPGGPMDEGPDLDGPPVDDDESDNSATVVGGLNDSVTLDLLA	378
QY	61	DFPKQGGVVKMKRQTQPMIHLYLDKETGPKKDATVSYEDPPTAKAAVEMPDGKDFQGS	120
DB	379	DFPKQGGVVKMKRQTQPMIHLYLDKETGPKKDATVSYEDPPTAKAAVEMPDGKDFQGS	438
QY	121	KLVSLARKKPPKPNNSRGGLPPREGGMPPLRGGGGGGGGGGGGGGGGGGGGGGGG	180
DB	439	KLVSLARKKPPKPNNSRGGLPPREGGMPPLRGGGGGGGGGGGGGGGGGGGGGGGG	498
QY	181	PRGPRGSRGNPSGGGVVORAGDMQCPNPGCGNPNAMPTGCGCAAPBEGFLPPPPPP	240
DB	499	PRGPRGSRGNPSGGGVVORAGDMQCPNPGCGNPNAMPTGCGCAAPBEGFLPPPPPP	558

QY 334 RDRPY 338
DB 400 RDRPY 404

RESULT 7

hypothetical protein C27H5.3 - *Caenorhabditis elegans*

C/Spectes: *Caenorhabditis elegans*

C/Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 09-Jul-2004

C/Accession: T15667

R/Author: A.

submitted to the EMBL Data Library, July 1995

A/Description: The sequence of *C. elegans* cosmid C27H5.

A/Reference number: Z18386

A/Accession: T15667

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 1-545 <PAU>

A/Cross-references: UNIPROT:Q18265; EMBL:U14635; NID:9540265; P1D:9540269; P1DN:AA046657

A/Experimental source: strain Bristol N2

C/Genetics:

A/Genes: C27H5.3

A/Intons: 38/2; 52/1; 87/3; 183/1; 202/2; 262/1; 308/2; 511/3

C/Superfamily: RNA-binding protein, EMS type; ribonucleoprotein repeat homology

Query Match 21.3%; Score 420.5; DB 2; Length 545;
Best Local Similarity 30.6%; Pred. No. 2e-19;
Matches 136; Conservative 29; Mismatches 95; Indels 184; Gaps 19;

QY 1 GGRGGM-----GSAGRGF-----NKGPG 20
DB 180 GGRGGYDGERGGSRNDGNSRQGGPGGGGGIYQIYVNCIFRTTSLFKLKNQY 239
QY 21 PMD-----EGPDLDGPPV-----DPDEDSD-----NS 43
DB 240 PTFEIEIQOKTLENYKNTCSYRGRPRD--GPPSGGGGGGGAASGNRBERGSDRYELKE 298
QY 44 AIYVQGLNDVTLDDLADFPKQGVVNMKRTQPMHIYLDKTEGPKGDAVTSYEDP 103
DB 299 TVFVQGISITANAYIADVSTGDIAXKND--GPRKITYDNTGPKBCMTITVDAS 356
QY 104 TAAAVEMFPGKDFOG--SKLAKYSLARKKPPMNSMKGLEPRRGRMPPLRGSGPG 160
DB 357 AAQQAATMTNKGPPPGSSSMSTSLAKFRADACGERG-----RGRGGFG 403
QY 161 GPGPMGRMGGRGDRGPPRGRSGRGNPSGGGVQHRAGDMQCPNPGCQNPFAMRT 220
DB 404 GRGPMGRMGGRGDRGPGGDRG--GGGGGRG----- 429
QY 221 ECHQCAKPKREGFLPPPPPPGDRGRGPGMGRGGLMDRGPGCMFRG--RGGDRG 279
DB 430 -----GPDGGRGGGGGFRGG-----DRGFRGGGRGGFRGGDRG 463
QY 280 GFRGG--RGMGRGGGGGRGGPGGPPPL-----MEQMG-----RR----- 315
DB 464 GFRGGDRGGDRGGRGR--GVGGGNANMEGRKNDWFCGCGSNFAPRECNQCAAPR 521
QY 316 -GGRGPGKMDKGEHNRDRRPY 318
DB 522 DGGSGGGGGGRGGRRGPGGDRRPY 545

RESULT 8

C96539

hypothetical protein P1413.10 [imported] - *Arabidopsis thaliana*

C/Spectes: *Arabidopsis thaliana* (mouse-ear crease)

C/Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 09-Jul-2004

C/Accession: C96539

R/Author: A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,

Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;

ansen, N.F.; Hughes, B.; Hultzar, L.

Nature 408, 816-820, 2000

A/Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.

C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luce, J.S.; Malt, R.; Marziani,

Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A/Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,

ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A/Title: Sequence and analysis of chromosome 1 of the plant *Arabidopsis*.

A/Reference number: A66141; PMID:21016719; PMID:11130712

A/Accession: C96539

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-211 <STO>

A/Cross-references: UNIPROT:Q95X47; GB:AE005173; NID:95734788; P1DN:AA050053.1; GSPDB:G

C/Genetics:

A/Genes: P1413.10

A/Map position: 1

Query Match 19.3%; Score 370.5; DB 2; Length 211;
Best Local Similarity 36.7%; Pred. No. 1.2e-16;
Matches 84; Conservative 21; Mismatches 91; Indels 33; Gaps 3;

QY 42 NSAIYQGLNDVTLDDLADFPKQGVVNMKRTQPMHIYLDKTEGPKGDAVTSYED 101
DB 7 NSVYVSNLPGLDENMLADYFTIGLKRDRKRTGPKWLYRDXETDEPKGDAVTSYED 66
QY 102 PTAAXAVEMFPGKDFOG--SKLAKYSLARKKPPMNSMKGLEPRRGRMPPLRGSGPG 161
DB 67 PTAAXAVEMFPGKDFOG--SKLAKYSLARKKPPMNSMKGLEPRRGRMPPLRGSGPG 114
QY 162 PGPMGRMGGRGDRGPPRGRSGRGNPSGGGVQHRAGDMQCPNPGCQNPFAMRT 221
DB 115 TNGGAGR-----GRGQDSAKXWQDDGMMCCNTSCTVNFAPRGV 156
QY 222 CHQCAKPKREGFLPPPPPPGDRGRGPGMGRGGLMDRGPGCMF 270
DB 157 CRRGCTARPAAGASGSGM--GAGRGGRGGGADGAPGKOPSGAPYGLP 202

RESULT 9

I50694 collagen alpha 1(III) chain - chicken (fragment)

C/Spectes: *Gallus gallus* (chicken)

C/Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 09-Jul-2004

C/Accession: I50694

R/Author: H.D.; Niu, Z.; Adams, S.L.

J. Biol. Chem. 269, 16443-16448, 1994

A/Title: An alternative transcript of the chick type III collagen gene that does not en

A/Reference number: A54041; PMID:94266842; PMID:8206952

A/Accession: I50694

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: mRNA

A/Residues: 1-886 <NAH>

A/Cross-references: UNIPROT:P12105; EMBL:U07973; NID:9520454; P1DN:AAA83407.1; P1D:95337

C/Genetics:

A/Genes: COL3A1

C/Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology

P;30-90/Domain: von Willebrand factor type C repeat homology <VWC>

Query Match 15.7%; Score 302; DB 2; Length 886;
Best Local Similarity 30.8%; Pred. No. 8.8e-12;
Matches 114; Conservative 20; Mismatches 104; Indels 132; Gaps 22;

QY 1 GGRGMSAGRGGRGGRG--GPMGDPDL-----GPPYDPEDSNNSAIYVQGLN 51
DB 584 GNEGADKNGERGGPGPGPTGPGAGKNGVGLGPPGPAAGPACDRGEPGSGPGLQGLP 643
QY 52 DSVTLDDLADFPKQGVVNMKRTQPMHIYLDKTEGPKGDAVTSYEDPTAAXAVEM 111
DB 644 GGP-----GPAENKRGEP-----GPKD----- 663
QY 112 FQKDFQSGSKLAKYSLARKKPPMNSMKGLEPRRGRMPPLRGSGPG-----GPG 163
DB 664 IGGPGRGPK-----GENGIRGERGPGPGPTGARGGPGAGSGAGKGP 709

QY 164 GPMGMRGGRG-----GDRGGRPPRRGSRGSRGNBSCGSGVQHRADNCRNPGC-----211
DB 710 GPPGAPGCGTGLPGLQGMPPGSRGASGSPGKDDGPEPGK-----ADGL--FGARGER 760
QY 212 GGNQFMARTECNQCKAPKPEGFLPPPPP-PGGDRGRGPGGPN---RGGRGLMDRG---264
DB 761 GN-----VGPICGPPRPPAPRPPDKETGTAGAPRPAASRGGRGGRGEG 803
QY 265 --GPGMFRG--GRGDRGFRGGRGMD--RGFRG---GRRGPG--GPPGLMEQMG 312
DB 804 LPPGAP-PPGAPGQNGEPG-KGBRGPPGLRGAPGAPGAPGCGPAPPPPP--GGVAK 859
QY 313 GRRGGRGGRG 322
DB 860 GERGSPGPG 869

RESULT 10
CGHUTL
collagen alpha 1(III) chain precursor - human
N/Alternate names: procollagen alpha 1(III) chain
C/Species: Homo sapiens (man)
C/Date: 24-Apr-1984 #sequence revision 01-Sep-1995 #text change 09-Jul-2004
C/Accession: S05272; S04642; PE0011; S01726; S04887; A90399; A94562; I51868; S59511; A90
R/Prockop, D.J.
Submitted to the EMBL Data Library, February 1989
A/Reference number: S05272
A/Accession: S05272
A/Status: preliminary
A/Molecule type: mRNA
A/Residues: 1-1240, 'V', 1242-1466 <PRC>
A/Cross-references: UNIPROT:P02461; EMBL:X14420; NID:G30057; PIDN:CAA32583.1; PID:G30058
R/Ala-Kokko, L.; Konturevskiy, S.; Baldwin, C.T.; Kivilahti, H.; Prockop, D.J.
Biochem. J. 260, 509-516, 1989
A/Title: Structure of cDNA clones coding for the entire prepro-alpha1(III) chain of human
collagen.
A/Reference number: S04642; NID:89350838; PMID:2764886
A/Accession: S04642
A/Molecule type: mRNA
A/Residues: 1-1196 <ALA>
A/Cross-references: GB:M26939; NID:G30057; PIDN:CAA32583.1; PID:G30058
R/Note: the complete sequence is not shown
R/Benson-Chanda, V.; Su, M.W.; Well, D.; Chu, M.L.; Ramirez, F.
Gene 78, 255-265, 1989
A/Title: Cloning and analysis of the 5' portion of the human type-III procollagen gene
A/Reference number: PE0011; NID:89378752; PMID:2777083
A/Accession: PE0011
A/Molecule type: DNA
A/Residues: 1-176 <BEN>
A/Cross-references: GB:M26939; NID:G180813; PIDN:AAA52040.1; PID:G180814
R/Toman, P.D.; Ricca, G.A.; de Crombrughe, B.
Nucleic Acids Res. 16, 7201, 1988
A/Title: Nucleotide sequence of a cDNA coding for the amino-terminal region of human pro
A/Reference number: S01726; NID:88303360; PMID:3405773
A/Accession: S01726
A/Molecule type: mRNA
A/Residues: 1-170 <OM>
A/Cross-references: EMBL:X07240; NID:G30060; PIDN:CAA30229.1; PID:G30061
R/Note: the authors translated the codon CAG for residue 154 as His
R/Janczko, R.A.; Ramirez, F.
Nucleic Acids Res. 17, 6742, 1989
A/Title: Nucleotide and amino acid sequences of the entire human alpha-1 (III) collagen.
A/Reference number: S04887; NID:89386015; PMID:2780304
A/Accession: S04887
A/Molecule type: mRNA
A/Residues: 149-163, 'G', 164-240, 'D', 242-471, 'D', 473-487, 'L', 489, 'S', 491-613, 'Y', 615-634,
A/Cross-references: EMBL:X15332; NID:G29545; PIDN:CAA33367.1; PID:G30045
R/Note: the authors' translation of residues 905-932 is inconsistent with the nucleotide
R/Seyer, J.M.; Kang, A.H.
Biochemistry 16, 1158-1164, 1977
A/Title: Covalent structure of collagen: amino acid sequence of cyanogen bromide peptide
A/Reference number: A90399; NID:77134724; PMID:557335

A/Accession: A90399
A/Molecule type: protein
A/Residues: 'V', 169-225, 229-232, 'P', 234-292, 'D', 294-398 <SEY1>
A/Experimental source: liver
A/Note: sequence corrected by A94562; attachment of 2-O-alpha-D-glucosyl-O-beta-D-galac
R/Seyer, J.M.
submitted to the Atlas, December 1977
A/Reference number: A94562
A/Accession: A94562
A/Molecule type: protein
A/Residues: 'V', 169-225, 229-277, 'A', 279-292, 'D', 294, 'S', 296-398 <SEY2>
A/Experimental source: liver
A/Note: author submitted corrections to A90399
R/Milewicz, D.M.; Wiltz, A.M.; Smith, A.C.; Manchester, D.K.; Waldstein, G.; Byers, P.H.
Am. J. Hum. Genet. 53, 62-70, 1993
A/Title: Parental somatic and germ-line mosaicism for a multilexon deletion with unusual
separating.
A/Reference number: I51868; NID:93304430; PMID:8317500
A/Accession: I51868
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 186-194 <MTL>
A/Cross-references: GB:S62925; NID:G386425; PIDN:AAJ3937.1; PID:G4261637
R/Chido, A.A.; Sillence, D.O.; Cole, W.G.; Bateman, J.F.
Biochem. J. 311, 939-943, 1995
A/Title: Abnormal type III collagen produced by an exon-17-skipping mutation of the COL
A/Reference number: S59511; NID:96067614; PMID:7487954
A/Accession: S59511
A/Molecule type: mRNA
A/Residues: 302-423 <CHI>
A/Cross-references: GB:S79877; NID:G1195576; PIDN:AA35615.1; PID:G1195577
R/Seyer, J.M.; Kang, A.H.
Biochemistry 17, 3404-3411, 1978
A/Title: Covalent structure of collagen: amino acid sequence of five consecutive CNBR p
A/Reference number: A90414; NID:79000343; PMID:687591
A/Accession: A90414
A/Molecule type: protein
A/Residues: 399-675, 'N', 677-727 <SEY3>
A/Experimental source: liver
R/Lee, B.; Vitale, B.; Superli-Furga, A.; Steinmann, B.; Ramirez, F.
J. Biol. Chem. 266, 5256-5259, 1991
A/Title: G to T transversion at position +5 of a splice donor site causes skipping of t
A/Reference number: I55349; NID:91161621; PMID:1672129
A/Accession: I55349
A/Status: translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 537-605 <LBS>
A/Cross-references: GB:M59312; NID:G180815; PIDN:AAA52041.1; PID:G180816
R/Seyer, J.M.; Mainardi, C.; Kang, A.H.
Biochemistry 19, 1583-1589, 1980
A/Title: Covalent structure of collagen: amino acid sequence of alpha1 (III)-CDS from t
A/Reference number: A90438; NID:80198282; PMID:6246925
A/Accession: A90438
A/Molecule type: protein
A/Residues: 728-895, 'A', 897-964 <SEY4>
A/Experimental source: liver
R/Cole, W.G.; Chido, A.A.; Lemande, S.R.; Janczko, R.; Ramirez, F.; Dahl, H.H.M.; Cha
J. Biol. Chem. 265, 17070-17077, 1990
A/Title: A base substitution at a splice site in the COL3A1 gene causes exon skipping a
A/Reference number: A38303; NID:91009133; PMID:2145268
A/Accession: A38303
A/Molecule type: mRNA
A/Residues: 861-1015 <COL>
A/Cross-references: GB:J05617; GB:M55603; GB:M59227; NID:G180878; PIDN:AA59383.1; PID:
A/Note: a mutant sequence with 942-977 spliced out from a patient with Ehlers-Danlos sy
R/Hancock, B.S.; Dalglish, R.
Nucleic Acids Res. 16, 2337, 1988
A/Title: Human pro alpha1(III) collagen: cDNA sequence for the 3' end.
A/Reference number: S02119; NID:88189827; PMID:3357782
A/Accession: S02119
A/Status: translation not shown
A/Molecule type: mRNA
A/Residues: 950-1018, 'Y', 1020-1183, 'S', 1185-1466 <MAN>

R: Seyer, J.M.; Kang, A.H.
Biochemistry 20, 2621-2627, 1981
A:Title: Covalent structure of collagen: amino acid sequence of alpha1 (III)-C9 from cy
A:Reference number: A90446; M0ID:81208139; PMID:7016180
A:Accession: A90446
A:Molecule type: protein
A:Residues: 965-979, 'N', 981-984, 'PS', 987, 'QN', 990-1096, 'P', 1098-1152, 'AT', 1155, 'S', 1157-
R: Loidl, H.R.; Brinker, J.M.; May, M.; Philhajanani, T.; Morrow, S.; Rosenbloom, J.; Mye
Nucleic Acids Res. 12, 9385-9394, 1984
A:Title: Molecular cloning and carboxyl-propeptide analysis of human type III procollage
Biochemistry 25, 1408-1413, 1986
A:Title: Human type III collagen gene expression is coordinately modulated with the type
A:Reference number: 152393; M0ID:86187804; PMID:3754462
A:Accession: 152393
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1161-1200 <MS>
A:Cross-references: GB:M13146; NID:g180415; PIDN:AA52003.1; PID:g180416
R: Remunet, B.S.; Camitzaro, L.A.; Seyer, J.M.; Myers, J.C.
Proc. Natl. Acad. Sci. U.S.A. 82, 3385-3389, 1985
A:Title: Human alpha 1(III) and alpha 2(V) procollagen genes are located on the long arm
A:Reference number: 159025; M0ID:85216505; PMID:3858826
A:Accession: 179359
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1165-1196 <MA>
A:Cross-references: GB:M11134; NID:g180417; PIDN:AA52004.1; PID:g180418
R: Chu, M.L.; Weil, D.; de Wet, W.; Bernard, M.; Sipola, M.; Ramirez, F.
J. Biol. Chem. 260, 4357-4363, 1985
A:Title: Isolation of cDNA and genomic clones encoding human pro-alpha1(III) collagen. H
A:Reference number: A92516; M0ID:85157600; PMID:2579949
A:Accession: A92516
A:Molecule type: DNA
A:Residues: 1176-1240, 'V', 1242-1356, 'P', 1358-1466 <CHU>
A:Cross-references: GB:M10615; GB:M10793; GB:M10794; GB:M10795; GB:M10796; GB:M10797; GB
A:Experimental source: liver
A:Note: the authors translated the codon TTC for residue 1057 as Tyr; the codons given f
action
C:Comment: Prolines and lysines at the third position of the tripeptide repeating unit
3-hydroxylated. About 15% of the lysines are 5-hydroxylated and some are subsequently C
C:Genetics:
A:Gene: GDB:COL3A1
A:Cross-references: GDB:118729; OMIM:120180
A:Map position: 2q31-2q31
A:Introns: 27/1, 94/3, 111/3, 149/3, 176/3, 554/3, 587/3, 1175/3, 1275/1, 1337/3, 1418/3
A:Note: the list of introns is incomplete; defects in this gene can result in Ehlers-D
C:Complex: type III collagen is a homotrimer of monomers initially linked by disulfide b
er of their length, is formed with desmosine cross-links made from lysine and allysine
C:Function:
A:Description: structural component of extracellular fibrous polymer that maintains inte
C:Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;
C:Keywords: coiled coil; Ehlers-Danlos syndrome; extracellular matrix; glycoprotein; hyd
P:1-23/Domains: signal sequence #status predicted <SIG>
P:124-153/Domains: amino-terminal propeptide #status predicted <PRO>
P:31-91/Domains: von Willebrand factor type C repeat homology <WVC>
P:154-1221/Product: collagen alpha 1(III) chain #status predicted <MAR>
P:154-167/Region: amino-terminal nonhelical telopeptide
P:168-1196/Region: helical
P:1091-1093/Region: cell attachment (R-G-D) motif
P:1197-1221/Region: carboxyl-terminal nonhelical telopeptide
P:1222-1466/Domains: carboxyl-terminal propeptide #status predicted <CPR>
P:1238-1466/Domains: fibrillar collagen carboxyl-terminal homology <FCC>
P:124/Modified site: pyrrolidone carboxylic acid (Gln) (in mature form) #status predicted
P:153-154/Cleavage site: Pro-Gln (procollagen N-endopeptide) #status predicted
P:154/Modified site: pyrrolidone carboxylic acid (Gln) (in mature form) #status predicted

[illegible]

Query Match	15.2%;	Score 293;	DB 2;	Length 1464;
Best Local Similarity	29.5%;	Pred. No. 5.1e-11;		
Matches 113;	Conservative 23;	Mismatches 117;	Indels 130;	Gaps 18;

RESULT 14

;Gene: CESP:Y66H1A.4

QY 143 REGGGMPPLDRGGPGCGGGPMGRMGRRGGDRGGFPFPRDPRGSKPNPSGGSNVQHANG 202
| | | | | | | | | | | | : |
Ddb 36 RGGTGGGGGFRGGRGGGGG---GFRRGGRGGRGG---GFRRGGRGFGGGGGRGGYIDG 87

RESULT 15

C:Accession: A90458; A90181; A02857
R:Highberger, J.H.; Corbett, C.; Dixit, S.N.; Yu, W.; Seyer, J.M.; Kang, A.H.; Gross, J

A:Title: Amino acid sequence of chick skin collagen alpha (I)-C88 and the complete primary structure of chick skin collagen alpha (I)-C88
A:Reference number: A90458; MWID:82231995; PMID:7093229

A;Molecule type: protein
A;Residues: 1-1036 <HIG>

A>Note: this is the latest in a series of papers from these workers elucidating the sequence of events leading to the development of the disease.

R.Eyre, D.R.; Glimcher, M.J.

A;Title: Evidence for a previously undetected sequence
A;Reference number: A90181; MUID:72243016; PMID:5047697

A;Molecule type: protein
A;Residues: 1037-1042 <EYR>

A>Note: residues 1037-1042 above correspond to the carboxyl end of the protein
C:Comment: Lysines at positions 103, 700, 934, and 946 above may be hydroxylated in some

Citation: Model of the proteins at the chondro-sarcoma, or endo-osteogenic sarcoma, of the bone.
CComment: Pro-1002 is the only 3-hydroxyproline and the only hydroxylated proline in p.
CSuperfamily: collagen alpha 1(I) chain; fibrillar collagen carboxy-terminal homology
CKeywords: collagen; extracellular matrix; glycoprotein; pyroglutamic acid; trimer;
F11Modified site: pyrrolidone carboxylic acid (Gln) #status experimental

Query Match	Score	DB 1	Length
14.8%	283.5	1042	

Best Local Similarity 29.6%; Pred. No. 1.5e-10;
Matches 110; Conservative 17; Mismatches 118; Indels 127; Gaps 18

4 GGMSSAGERGGFNKGGPMDEGPDLDLGPVPDPPDESDNSAIYVQGLNDSVTLDDLDLDF 63

64 KQCGVVKNNKRTGQBMHILYLDKETGKPKGDATVSIEDPPTAKAAVEMFDFGKDFQGSXLK 123

Db 60 -PAGEGXNDGD-----EAGKP---GRPGRGPPRGGA-----RGLPGT 96

Db 97 AGL-----PGKKGHRFPGLDGAQKQPGPAPKPGEPSPSPGENGAP--CGMGPRGLPGERGRP 151

180 PPRGDRGSRGNPSSGGGNVQHRADGNOCPNPGCGNQNFAMRTECNQCKAPKPEGFLPPPP 231

Db	152	GPSPGAPAGANDAPG----	AAEP----	PGP-----	-----	TGPAG--	PPGFP	184
Qy	240	PPGDDRRGRGPPGMMRGGRGGLMDRGGPGGMFRGGRGDRG-						279
Db	185	GAAGAKKETPPQAGRGSEBPPQSGRSEBPPGAPAGAPAGNPGADGPGAKGATGAPGIA						244
Qy	280	---GFRGGRGMD-----	RGFRG-----	-----	GGRRGPG--	-----	GPPGLMEQ	310
Db	245	GAPGPPGAPGSPGPPQPSGAPGPPKNGSGSEBAPGNKGTGAKGEPGAPGVQGPSPGAGEB						304
Qy	311	MGRRGGRGGGPG	322					
Db	305	--GKRGARSEBPG	314					

Search completed: February 18, 2005, 15:32:15
Job time : 49 secs

This Page Blank (uspto)

Query Match	100.0%	Score 1922;	DB 2;	Length 600;
Best Local Similarity	100.0%	Pred. No. 3.6e-103;		
Matches 338;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0
1	GGGGGMSAGERGCGNNKPCGPMDEGPDLDLGPVPDPDESDNSAIYVGLANDSVTLDDL	60		
263	GGGGGMSAGERGCGNNKPCGPMDEGPDLDLGPVPDPDESDNSAIYVGLANDSVTLDDL	322		
61	DFFKCGVVKNNKRTQPMTHIYLDLKEGPKGDAIVSYEDPPTAKAAVEMFDGKDFGS	120		
323	DFFKCGVVKNNKRTQPMTHIYLDLKEGPKGDAIVSYEDPPTAKAAVEMFDGKDFGS	382		
121	KLKVSLLAKKKPPMNSMRGSLPPREGRGMPPLRGGPGGGGGGGMGRGGRGGRGGR	180		
363	KLKVSLLAKKKPPMNSMRGSLPPREGRGMPPLRGGPGGGGGGGMGRGGRGGRGGR	442		
161	PRGPRGSRNPGSGGNVOHRAADWOCNPNCGCNQNPAMWTECNQCKAPKPEGFLPPPPP	240		
443	PRGPRGSRNPGSGGNVOHRAADWOCNPNCGCNQNPAMWTECNQCKAPKPEGFLPPPPP	502		
241	PGDSDGRGPGGMRGGRGGLMDRGGPGFRGGGGRGGRGGRGGRGGRGGRGGR	300		
503	PGDSDGRGPGGMRGGRGGLMDRGGPGFRGGGGRGGRGGRGGRGGRGGRGGR	562		
301	GGPPGPPLMEQMGRRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGR	338		
563	GGPPGPPLMEQMGRRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGR	600		

RN SEQUENCE FROM N.A.
 RP TISSUE=Placenta;
 RC MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.2426038992;
 RA Straubeberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shemen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Bucoff K.H., Scheater C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA D'Antonio L., Maruina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stedman M., Soares M.B., Bonaldi M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carinici P., Prange C.J.,
 RA Rata S.S., Loguigliano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.V., Malek J.A., Gamaralle P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Ketterman M., Madan A., Rodrigues S., Sanchez A.,
 RA Whiting R.W., Madan A., Young A.C., Shevchenko V., Bouffard G.G.,
 RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Buttefield J.S.N., Krzywicki M.I., Skalski S., Smalies D.E.,
 RA Schmechel A., Schein J.E., Jones S.J.M., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length human
 RT and mouse cDNA sequences."
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RN
 RN SEQUENCE OF 1-345 FROM N.A.
 RX MEDLINE=97115101; PubMed=89756699; DOI=10.1006/geno.1996.0625;
 RA Zucman-Rossi J., Legoux P., Thomas G.;
 RT "Identification of new members of the Gas2 and Ras families in the
 RL 22q12 chromosome region."
 RL Genomics 38:247-254(1996).
 RN
 RN SEQUENCE OF 241-266 FROM N.A.
 RC TISSUE=Placenta;
 RX MEDLINE=95352541; PubMed=7542907;
 RA Bhagirath T., Abe S., Nojima T., Yoshida M.C.;
 RT "Molecular analysis of a t(11;22) translocation junction in a case of
 RT Ewing's sarcoma."
 RL Genes Chromosomes Cancer 13:126-132(1995).
 RN
 RN PHOSPHORYLATION SITE SER-266, AND PARTIAL SEQUENCE.
 RP MEDLINE=98001723; PubMed=9341188; DOI=10.1074/jbc.272.4.73269;
 RA Deloume J.C., Pritchard L., Delatour O., Storm D.R.;
 RT "The pro-oncogene EWS binds calmodulin and is phosphorylated by
 RT protein kinase C through an IQ domain."
 RL J. Biol. Chem. 272:27369-27377(1997).
 RN
 RN SEQUENCE OF 128-158; 233-247; 268-324; 334-364; 393-439; 447-518 AND
 RP 551-641, METHYLATION OF ARGININES, AND MASS SPECTROMETRY.
 RX MEDLINE=21276345; PubMed=11278906; DOI=10.1074/jbc.M011446200;
 RA Beljaneckaya L.V., Gehrig P.M., Gehrig H.;
 RT "Exposure on cell surface and extensive arginine methylation of Ewing
 RT sarcoma (EWS) protein."
 RL J. Biol. Chem. 276:18681-18687(2001).
 RN
 RN ALTERNATIVE SPLICING, AND RNA-BINDING.
 RP MEDLINE=94366763; PubMed=8084618;
 RX Ohno T., Ouchida M., Lee L., Gatalica Z., Rao V.N., Reddy E.S.P.;
 RA "The EWS gene, involved in Ewing family of tumors, malignant melanoma
 RT of soft parts and desmoplastic small round cell tumors, codes for an
 RT RNA binding protein with novel regulatory domains."
 RL Oncogene 9:3087-3097(1994).
 RN
 RN INTERACTION WITH SP1.
 RP MEDLINE=98325009; PubMed=9660765; DOI=10.1074/jbc.273.29.18086;
 RA Zhang D., Paley A.J., Childs G.;
 RT "The transcriptional repressor ZNF1 interacts with and modulates the
 RT ability of EWS to activate transcription."
 RL J. Biol. Chem. 273:18086-18091(1998).
 RN
 RN CHARACTERIZATION.
 RP MEDLINE=20390060; PubMed=10767297; DOI=10.1074/jbc.M002961200;
 RX Li K.K.C., Lee K.A.W.;

[1] SEQUENCE FROM N.A.
TISUSe=Placentat;
MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
Straussberg R.L., Retingold E.A., Grouse L.H., Derge J.G.,
Klausner R.D., Collins F.S., Wagner L., Sherman C.M., Schuler G.D.,
Altchul S.F., Zeeberg B., Butelov K.H., Scheffer C.F., Bat N.K.,
Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
Dachancko L., Marisina K., Farmer A.A., Rubin G.M., Hong L.,
Stapleton M., Soares M.B., Bonaldi M.F., Casavant T.L., Scheetz T.E.,
Brownstein M.J., Udell T.B., Toshiyuki S., Carinci P., Prange C.,
Raha S.S., Loguolino N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
Bosak S.A., McEwan K.J., McKernan K.J., Malek J.A., Gumarathne P.H.,
Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Huliy S.W.,
Falla J., Helton E., Kettman M., Madan A., Lu X., Gibbs R.A.,
Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
Butterfield Y.S.N., Krzywnicki M.I., Skalska U., Smalins D.E.,
Schneerch A., Schein J.E., Jones S.J.M., Marra M.A.;
"Generation and initial analysis of more than 15,000 full-length human
and mouse cDNA sequences";
Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).

[5] SEQUENCE OF 1-345 FROM N.A.
MEDLINE=97131501; PubMed=975699; DOI=10.1006/geno.1996.0625;
Zucman-Rossi J., Legoux P., Thomas G.;
"Identification of new members of the Gatz and Ras families in the
22q12 chromosome region";
Genomics 38:247-254(1996).

[6] SEQUENCE OF 241-268 FROM N.A.
TISUSe=Placentat;
MEDLINE=95352541; PubMed=7542907;
Bhagirath T., Abe S., Nojima T., Yoshida M.C.;
"Molecular analysis of a t(11;22) translocation junction in a case of
Ewing's sarcoma";
Genes Chromosomes Cancer 13:126-132(1995).

[7] PHOSPHORYLATION SITE SER-266, AND PARTIAL SEQUENCE.
MEDLINE=98001723; PubMed=3341188; DOI=10.1074/jbc.272.43.27369;
Delouche J.C., Pichard L., Delatour O., Storm D.R.;
"The prooncogene EWS binds calmodulin and is phosphorylated by
protein kinase C through an IQ domain";
J. Biol. Chem. 272:27369-27377(1997).

[8] SEQUENCE OF 128-158; 233-247; 268-324; 334-364; 393-439; 447-518 AND
551-641, METHYLATION OF ARGININES, AND MASS SPECTROMETRY.
MEDLINE=11276345; PubMed=11278966; DOI=10.1074/jbc.M011446200;
Belyankaya L.L., Gehrig P.M., Gehrig H.;
"Exposure on cell surface and extensive arginine methylation of Ewing
sarcoma (EWS) protein";
J. Biol. Chem. 276:18681-18687(2001).

[9] ALTERNATIVE SPLICING, AND RNA-BINDING.
MEDLINE=94366763; PubMed=8084618;
Ohno T., Ouchida M., Ise L., Catalica Z., Rao V.N., Reddy E.S.P.;
"The EWS gene, involved in Ewing family of tumors, malignant melanoma
of soft parts and desmoplastic small round cell tumors, codes for an
RNA binding protein with novel regulatory domains";
Oncogene 9:3087-3097(1994).

[10] INTERACTION WITH SP1.
MEDLINE=98825009; PubMed=9660765; DOI=10.1074/jbc.273.29.18086;
Zhang D., Paley A.U., Childs G.;
"The transcriptional repressor ZEM1 interacts with and modulates the
ability of EWS to activate transcription";
J. Biol. Chem. 273:18086-18091(1998).

[11] CHARACTERIZATION
MEDLINE=20390060; PubMed=10767297; DOI=10.1074/jbc.M002961200;
MEDLINE=20390060; PubMed=10767297; DOI=10.1074/jbc.M002961200;
J.K.K.C., Lee K.A.W.;

RT "Transcriptional activation by the Ewing's sarcoma (EWS) oncogene can
RT be cis-repressed by the EWS RNA-binding domain.",
RT J. Biol. Chem. 275:23053-23058(2000).
CC -1- FUNCTION: Might normally function as a repressor. EWS-fusion-
CC proteins (EFPs) may play a role in the tumorigenic process. They
CC may disturb gene expression by mimicking, or interfering with the
CC normal function of CTD-Polr1 within the transcription initiation
CC complex. They may also contribute to an aberrant activation of the
CC fusion protein target genes.
CC -1- SUBUNIT: Binds POLR2C, SFI, calmodulin and RNA. Interacts with
CC PTK2B/PKX2.
CC -1- SUBCELLULAR LOCATION: Nuclear, cytoplasmic and on cell surface.
CC Relocates from cytoplasm to ribosomes upon PRK2/PKX2 activation.
CC -1- ALTERNATIVE PRODUCTS:
CC Name=EWS;
CC IsoId=Q01844-1; Sequence=Displayed;
CC Name=EWS-B;
CC IsoId=Q01844-2; Sequence=VSP_005793;
CC Note=No experimental confirmation available;
CC -1- TISSUE SPECIFICITY: Ubiquitous.
CC -1- DOMAIN: EWS activation domain (EAD) functions as a potent
CC activation domain in EFPs. EWSR1 binds POLR2C but not POLR2B or
CC POLR2G, whereas the isolated EAD binds POLR2E and POLR2G but not
CC POLR2C. Cts-linked RNA-binding domain (RBD) can strongly and
CC specifically repress trans-activation by the EAD.
CC -1- PTM: Phosphorylated; calmodulin-binding inhibits phosphorylation
CC of Ser-266.
CC -1- DISEASE: Ewing's sarcoma is characterized by chromosomal
CC translocations t(11;22)(q24;q12) which involves EWSR1 and FLI1,
CC t(7;22)(p22;q12) which involves EWSR1 and ETV1, t(21;22)(q22;q12)
CC which involves EWSR1 and NR4A3.
CC -1- DISEASE: Involved in desmoplastic small round cell tumor (DSRCT)
CC through a chromosomal translocation t(11;22)(p13;q12) that
CC involves EWSR1 and WT1.
CC -1- DISEASE: Malignant melanoma of soft parts (MMSP), also known as
CC soft tissue clear cell sarcoma, is a rare tumor developing in
CC tendons and aponeuroses. It is associated with chromosomal
CC translocation t(12;22)(q13;q12) involving EWSR1 and ATF-1.
CC -1- DISEASE: Involved in small round cell sarcoma through a
CC chromosomal translocation t(1;22)(p36.1;q12) that involves EWSR1
CC and ZNF278.
CC -1- MISCELLANEOUS: EFPs arise due to chromosomal translocations in
CC which EWSR1 is fused to a variety of cellular transcription
CC factors. EFPs are very potent transcriptional activators dependent
CC on the EAD and a C-terminal DNA-binding domain contributed by the
CC fusion partner. The spectrum of malignancies associated with EFPs
CC are thought to arise via EFP-induced transcriptional deregulation,
CC with the tumor phenotype specified by the EWSR1 fusion partner and
CC cell type. Transcriptional repression by the EWSR1 fusion partner and
CC factor beta type II receptor (TRGF beta RII) is an important target
CC of the EWS-FLI1, EWS-ERG, or EWS-ETV1 oncogene.
CC -1- MISCELLANEOUS: Binds calmodulin in the presence, but not in the
CC absence, of calcium ion.
CC -1- SIMILARITY: Belongs to the RNP TET family.
CC -1- SIMILARITY: Contains 1 IQ domain.
CC -1- SIMILARITY: Contains 1 RanBP2-type zinc finger.
CC -1- SIMILARITY: Contains 1 RNA recognition motif (RNM) domain.
CC -1- DATABASE: NAME=Atlas Genet. Cytogenet. Oncol. Hematol.;
CC WWW="http://www.infobiogen.fr/services/chronocancer/Gene/EWSR11085.html".
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X66899; CAA47350.1; -
CC EMBL; X72990; CAA51489.1; -
CC EMBL; X72991; CAA51489.1; JOINED.

DR EMBL; X72992; CAA51489.1; JOINED.
DR EMBL; X72993; CAA51489.1; JOINED.
DR EMBL; X72994; CAA51489.1; JOINED.
DR EMBL; X72995; CAA51489.1; JOINED.
DR EMBL; X72996; CAA51489.1; JOINED.
DR EMBL; X72997; CAA51489.1; JOINED.
DR EMBL; X72998; CAA51489.1; JOINED.
DR EMBL; X72999; CAA51489.1; JOINED.
DR EMBL; X73000; CAA51489.1; JOINED.
DR EMBL; X73001; CAA51489.1; JOINED.
DR EMBL; X73002; CAA51489.1; JOINED.
DR EMBL; X73003; CAA51489.1; JOINED.
DR EMBL; X73004; CAA51489.1; JOINED.
DR EMBL; BC004817; AA04817.1; -
DR EMBL; Y08806; CAA70044.1; ALT_INIT.
DR EMBL; AB016435; BAA31990.1; -
DR EMBL; Y07848; CAA69177.1; -
DR F1R; A49358; A49358.
DR HSSP; Q95218; 1N0Z.
DR GeneW; HGNC:3508; EWSR1.
DR H-INVDB; HIX0016349; -
DR MIM; 133450; -

Query Match 100.0%; Score 1922; DB 1; Length 656;
Best Local Similarity 100.0%; Pred. No. 3.9e-103;
Matches 338; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGRGWSAGERGFNKPQPGMDDEGPDLDLGPVDPDEBDSNSAIVVQGLNSVTLLDILA 60
DB GGRGWSAGERGFNKPQPGMDDEGPDLDLGPVDPDEBDSNSAIVVQGLNSVTLLDILA 378
QY 61 DFFKQGVVYVNRKTRQPMIHLYLDETKGPKKDATVSIJBDPTTAAVWFPGKPOGS 120
DB DFFKQGVVYVNRKTRQPMIHLYLDETKGPKKDATVSIJBDPTTAAVWFPGKPOGS 438
QY 121 KLVSLARKKPPNNSRGGLPPEBGRGMPPLRGGGGGGGPGPMRGGRGDRGGFP 180
DB KLVSLARKKPPNNSRGGLPPEBGRGMPPLRGGGGGGGPGPMRGGRGDRGGFP 498
QY 439 KLVSLARKKPPNNSRGGLPPEBGRGMPPLRGGGGGGGPGPMRGGRGDRGGFP 498
DB KLVSLARKKPPNNSRGGLPPEBGRGMPPLRGGGGGGGPGPMRGGRGDRGGFP 558
QY 181 PGRPRSRNPGSGGVVQHRAGDWCQPNPDCGNQNFAMWTECNQCAKPPREGFLPPFPFP 240
DB PGRPRSRNPGSGGVVQHRAGDWCQPNPDCGNQNFAMWTECNQCAKPPREGFLPPFPFP 558
QY 499 PGRPRSRNPGSGGVVQHRAGDWCQPNPDCGNQNFAMWTECNQCAKPPREGFLPPFPFP 558
DB PGRPRSRNPGSGGVVQHRAGDWCQPNPDCGNQNFAMWTECNQCAKPPREGFLPPFPFP 618
QY 241 PGCDRGRGPGGMRGGRGGLMDRGFGGMPFRGGRGGDRGFRGGRGMDRGFGGRRGPG 300
DB PGCDRGRGPGGMRGGRGGLMDRGFGGMPFRGGRGGDRGFRGGRGMDRGFGGRRGPG 618
QY 559 PGCDRGRGPGGMRGGRGGLMDRGFGGMPFRGGRGGDRGFRGGRGMDRGFGGRRGPG 618
DB PGCDRGRGPGGMRGGRGGLMDRGFGGMPFRGGRGGDRGFRGGRGMDRGFGGRRGPG 618
QY 301 GGPPLMEOMGGRGGRGGRGPGGPMGXGHRORRRDPY 338
DB GGPPLMEOMGGRGGRGGRGPGGPMGXGHRORRRDPY 656
QY 619 GGPPLMEOMGGRGGRGGRGPGGPMGXGHRORRRDPY 656
DB GGPPLMEOMGGRGGRGGRGPGGPMGXGHRORRRDPY 656

RESULT 3
Q96FE8 PRELIMINARY; PRT; 655 AA.
AC Q96FE8;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 25-OCT-2004 (TREMBLrel. 28, Last annotation update)
GN Ewing sarcoma breakpoint region 1, isoform EWS (EWSR1 protein).
GN Name=EWSR1;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Lymph, and Skin;
RC MEDLINE=2388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Klausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Strausberg R.D., Collins F.S., Wagner L., Shennan C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,

RA Datchenko L., Marusina K.B., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Ustin T.B., Toshynski S., Carlini P., Prange C.J.,
Rana S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
Rosa S.A., McEwen P.J., McKernan K.J., Malek J.C., Gunataratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahney J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shcherchenko Y., Bouffard G.G.,
RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Buterfield J.S.,
RA Krzywinski M.I., Skalska U., Smallus D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Maxam M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).

RN (2)
RP SEQUENCE FROM N.A.
RC TISSUE=Lymph;
RA Straubeberg R.;
RL Submitted (JUL-2001) to the EMBL/GenBank/DDBJ databases.

RN (3)
RP SEQUENCE FROM N.A.
RC TISSUE=Skin;
RA Straubeberg R.;
RL Submitted (JUN-2004) to the EMBL/GenBank/DDBJ databases.

RN (4)
RP SEQUENCE FROM N.A.
RA Colline J.E., Wright C.L., Edwards C.A., Davis M.P., Grisham J.A.,
RA Cole C.G., Goward M.E., Aguado B., Malliya M., Mokrab Y., Huckle E.J.,
Beare D.M., Dunham I.;
RL Submitted (MAY-2004) to the EMBL/GenBank/DDBJ databases.

DR EMBL; BC011048; AAH11046.1; -
DR EMBL; BC072442; AAH72442.1; -
DR EMBL; CR456490; CACG30376.1; -
DR HSSP; O95218; INOZ.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003676; F:catalytic acid binding; IEA.
DR GO; GO:0008270; F:zinc ion binding; IEA.
DR InterPro; IPRO11368; RNA-binding_EWS.
DR InterPro; IPRO00504; RNA_rec_mot.
DR InterPro; IPRO01876; Znf_RangBP.
DR Pfam; PF00076; RRM_1; 1.
DR Pfam; PF00641; ZF-RANBP_1.
DR PIRSF; PIRSF002101; RNA-binding_EWS; 1.
DR SMART; SM00360; RRM; 1.
DR SMART; SM00547; ZNF_RBZ; 1.
DR PROSITE; PS0102; RRM; 1.
DR PROSITE; PS01358; ZF_RANBP2_1; 1.
DR PROSITE; PS01199; ZF_RANBP2_2; 1.
SO SEQUENCE 655 AA; 68391 MW; B539ED1B98C601ED CRC64;

Query Match 99.2%; Score 1907.5; DB 2; Length 655;
Best Local Similarity 99.7%; Pred. No. 2.7e-102;
Matches 337; Conservative 0; Mismatches 0; Indels 1; Gaps 1

Gy 1 GGRGMGSAGERGNNKPGEPMDDEGPDLGLGPVPDDESDNSAIYVGQGLNDSTVLLDILA 60
Db 319 GGRGGMG-AGERGFNKKPCGPMDEGPDLDLGPPVDDESDNSAIYVGQLNDSTVLLDILA 377
Gy 61 DPFKCGGVYKNNKRTRGQPPIHIYYLDKETGKPRFGDAIVSYEDPPTAKAAVENWTDGDQFQS 120
Db 378 DFFKCQGVYKNNKRTRGQPPIHIYYLDKETGKPRFGDAIVSYEDPPTAKAAVENWTDGDQFQS 437
Gy 121 KLKVLGSLARKKPENMSNRGGLPRERGRMPPLRGSGGPGGPGGPMGRGGRGGRGDPFP 180
Db 438 KLKVLGSLARKKPENMSNRGGLPRERGRMPPLRGSGGPGGPGGPMGRGGRGGRGDPFP 497
Gy 181 PRGPRGSRGNPSGGGAVNOHRADWQCNPNGCNONFAMRTCCNQCKAPKPBEGFLPPPFP 240
Db 498 PRGPRGSRGNPSGGGAVNOHRADWQCNPNGCNONFAMRTCCNQCKAPKPBEGFLPPPFP 557
Gy 241 PGGBRGRRGPGGMRGGRGGLMDRGCPGMFRGGRGGRGDRGFRGGRGMRDRCFTGGRRGSP 300

D6		558	P6GDNRGCGP	GCGMRGGRGGGLMDRGAPCGCMFGGCGJRRGGRGGRGMDRGGFCGGRGCGP	61.7
Oy	301	GGPPGPLMEOWGRRGRRGRRGPCKDXXGHRERRDRPY	338		
D6	618	GGPPGPLMEOWGRRGRRGRRGPCKDXXGHRERRDRPY	655		
<hr/>					
RESULT 4					
O96MX4		PRELIMINARY;	PRT;	661 AA.	
ID	O96MX4				
AC	O96MX4				
DT	01-DEC-2001	(TEMBLrel. 19, Created)			
DT	01-DEC-2001	(TEMBLrel. 19, Last sequence update)			
DT	01-MAR-2004	(TEMBLrel. 26, Last annotation update)			
DE	Hypothetical protein FLJ31747.				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
OX	NCBI_TaxId=9606;				
RP		[1]			
RX	SEQUENCE FROM N.A.				
RX	PubMed=14702039; DOI=10.1038/ngl285;				
RX	Ota T., Suzuki Y., Nishikawa T., Otsuki T., Sugiyama T., Irie R.,				
RX	Kamakura A., Hayashi K., Sato H., Nagai K., Kimura K., Makita H.,				
RX	Selme M., Obayashi M., Nishi T., Shibahara T., Tanaka T., Ishii S.,				
RX	Yamamoto J., Saito K., Kawai Y., Isono Y., Nakamura Y., Nagahari K.,				
RX	Murakami K., Yasuda T., Iwayanagi T., Nagatsuma M., Shiratori A.,				
RX	Sudo H., Hogoishi T., Kaku Y., Kodaira H., Kondoh H., Sugawara M.,				
RX	Takahashi M., Kanda K., Yokoi T., Futuya T., Tokikawa E., Omura Y.,				
RX	Abe K., Kamihara K., Katsuma N., Sato K., Tanikawa M., Yamazaki M.,				
RX	Ninomiya K., Ishibashi T., Yamashita H., Murakawa K., Fujimori K.,				
RX	Tanai H., Kimura M., Watanabe M., Hirooka S., Chiba Y., Iehida S.,				
RX	Ono Y., Takiguchi S., Watanabe S., Yoshida M., Hotuta T., Kusano J.,				
RX	Kanehori K., Takahashi-Fujii A., Hara H., Tanase T., Nomura Y.,				
RX	Tojima S., Komai F., Hara R., Takeuchi K., Arita M., Imose N.,				
RX	Maeshino K., Yuuki H., Oshima A., Sasaki N., Aotuka S.,				
RX	Yoshikawa Y., Matunawa H., Ichihara T., Shiohata N., Sano S.,				
RX	Moriya S., Momiyama H., Satoh N., Takami S., Terashima Y., Suzuki O.,				
RX	Nakagawa S., Senoh A., Mizoguchi H., Goto Y., Shimizu F., Wakebe H.,				
RX	Hishigaki H., Watanabe T., Sugiyama A., Takemoto M., Kawakami B.,				
RX	Yamazaki M., Watanabe K., Kumagai A., Itakura S., Fukuzumi Y.,				
RX	Fujimori Y., Komiyama M., Tashiro H., Tanigami A., Fujiwara T.,				
RX	Ono T., Yamada K., Fujii Y., Ozaki K., Hirao M., Ohmori Y.,				
RX	Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,				
RX	Ohtani R., Kawakami T., Noguchi S., Itoh T., Shigetake S., Senba T.,				
RX	Matsunuma K., Nakajima Y., Mizuno T., Morinaga M., Sasaki M.,				
RX	Tozashiki T., Oyama M., Hata H., Watanabe M., Komatsu T.,				
RX	Mitsushima-Sugano J., Satoh T., Shirai Y., Takahashi Y., Nakagawa K.,				
RX	Okumura K., Nagase T., Nomura N., Kikuchi H., Masuo Y., Yanashita R.,				
RX	Nakai K., Yada T., Nakamura Y., Ohara O., Isogai T., Sugano S.,				
RT	"Complete sequencing and characterization of 21,243 full-length human				
RT	cDNAs."				
RL	Nat. Genet. 36:40-45(2004).				
EMBL	AK056309; BAB7145.1; -				
DR	HSSP; O95218; INOZ.				
DR	GO; GO:0005634; C:nucleus; IEA.				
DR	GO; GO:0003676; P:nucleic acid binding; IEA.				
DR	GO; GO:0008270; P:zinc ion binding; IEA.				
DR	InterPro; IPRO11368; RNA-binding_EWS.				
DR	InterPro; IPRO00504; RNA_rec_mot.				
DR	InterPro; IPRO01876; Znf_RangDP.				
DR	Pfam; PF00076; RFM_1; 1.				
DR	Pfam; PF00641; zf-RandP; 1.				
DR	PIRSF; PIRSF002101; RNA-binding_EWS; 1.				
DR	SMART; SM00360; RMW; 1.				
DR	SMART; SM00547; ZNF_RBZ; 1.				
DR	PROSITE; PS50102; RFM; 1.				
DR	PROSITE; PS01358; ZF_RANBP2_1; 1.				
DR	PROSITE; PS01399; ZF_RANBP2_2; 1.				
SG	SEQUENCE 661 AA; 6696 MW; 5F84F52FDCCD51269 CRC64;				

Best Local Similarity 99.7%; Pred. No. 2,7e-102; Matches 337; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1 GGRGMSAGRGGFPKPGMDGPDLDGPVDPEDSDNSAIYVQGLNDSVTLDDL 60
 DB 325 GGRGMSAGRGGFPKPGMDGPDLDGPVDPEDSDNSAIYVQGLNDSVTLDDL 383
 QY 61 DFFKQGVVMMNRRTGQPMHIYLDKRTGPKGDATVSYEDPTAKAAVWFDKDFQGS 120
 DB 384 DFFKQGVVMMNRRTGQPMHIYLDKRTGPKGDATVSYEDPTAKAAVWFDKDFQGS 443
 QY 121 KLVSLARKKPPMNSMRGGLPPREGMPPLRLGPGPGPGPGMGRMGGRGDRGFP 180
 DB 444 KLVSLARKKPPMNSMRGGLPPREGMPPLRLGPGPGPGPGMGRMGGRGDRGFP 503
 QY 181 PRGPRGSRGNPSGGGNVQHRAGDMQCPNPGCQGNFAMRTGKQCAKPKPEGFLPPFP 240
 DB 504 PRGPRGSRGNPSGGGNVQHRAGDMQCPNPGCQGNFAMRTGKQCAKPKPEGFLPPFP 563
 QY 241 PGDGRGGRGPGMRGGRGGLMDRGPGGMRGGRGGRGGRGGRGGRGGRG 300
 DB 564 PGDGRGGRGPGMRGGRGGLMDRGPGGMRGGRGGRGGRGGRGGRGGRG 623
 QY 301 GGPPLMEQMGRRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRG 338
 DB 624 GGPPLMEQMGRRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRG 661

RESULT 5
 Q6NVA3 PRELIMINARY; PRT; 656 AA.

AC 06NVA3;
 DT 05-JUL-2004 (TEMBLrel. 27, Created)
 DT 05-JUL-2004 (TEMBLrel. 27, Last sequence update)
 DE Ewing sarcoma breakpoint region 1.
 GN Name=Ewari;
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6; TISSUE=Brain;
 RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603999;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klauener R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altshul S.F., Zeeberg B., Buettow K.H., Schaefer C.F., Hsieh F.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusik K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Uedin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Morley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
 RA Blakeley R.W., Touchman J.W., Green E.D., Dickson G.G.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
 RA Krzywicki M.I., Skalka U., Smalton D.E., Schmechel A., Schein J.E.,
 RA Jones S.J., Matra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length human
 and mouse cDNA sequences";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6; TISSUE=Brain;
 RA Director MGC Project;
 DR EMBL; BC068226; AAH68226.1;
 DR GO; GO:0005634; C:nucleus; IDA.
 DR InterPro; IPR011368; RNA-binding_EWS.

DR InterPro; IPR000504; RNA_rec_mol.
 DR InterPro; IPR001876; Znf_RanBDP.
 DR Pfam; PF00076; RRM_1; 1.
 DR Pfam; PF00641; zf_RanBP; 1.
 DR PIRSF; PIRSF02101; RNA-binding_Ews; 1.
 DR SMART; SM00360; RRM; 1.
 DR SMART; SM00547; Znf_RBZ; 1.
 DR PROSITE; PS0102; RRM; 1.
 DR PROSITE; PS01358; zf_RanBP2_1; 1.
 DR PROSITE; PS01399; zf_RanBP2_2; 1.
 SQ SEQUENCE 656 AA; 68549 MW; 5DF018A22C8C10B CRC64;

Query Match 99.0%; Score 1903; DB 2; Length 656;
 Best Local Similarity 98.5%; Pred. No. 4.9e-102;
 Matches 337; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 1 GGRGMSAGRGGFPKPGMDGPDLDGPVDPEDSDNSAIYVQGLNDSVTLDDL 60
 DB 319 GGRGMSAGRGGFPKPGMDGPDLDGPVDPEDSDNSAIYVQGLNDSVTLDDL 378
 QY 61 DFFKQGVVMMNRRTGQPMHIYLDKRTGPKGDATVSYEDPTAKAAVWFDKDFQGS 120
 DB 379 DFFKQGVVMMNRRTGQPMHIYLDKRTGPKGDATVSYEDPTAKAAVWFDKDFQGS 438
 QY 121 KLVSLARKKPPMNSMRGGLPPREGMPPLRLGPGPGPGPGMGRMGGRGDRGFP 180
 DB 439 KLVSLARKKPPMNSMRGGLPPREGMPPLRLGPGPGPGPGMGRMGGRGDRGFP 498
 QY 181 PRGPRGSRGNPSGGGNVQHRAGDMQCPNPGCQGNFAMRTGKQCAKPKPEGFLPPFP 240
 DB 499 PRGPRGSRGNPSGGGNVQHRAGDMQCPNPGCQGNFAMRTGKQCAKPKPEGFLPPFP 558
 QY 241 PGDGRGGRGPGMRGGRGGLMDRGPGGMRGGRGGRGGRGGRGGRGGRG 300
 DB 559 PGDGRGGRGPGMRGGRGGLMDRGPGGMRGGRGGRGGRGGRGGRGGRG 618
 QY 301 GGPPLMEQMGRRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRG 338
 DB 619 GGPPLMEQMGRRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRG 656

RESULT 6
 EWS_MOUSE
 ID EWS_MOUSE
 AC 061545;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 25-OCT-2004 (Rel. 45, Last annotation update)
 DE RNA-binding protein EWS.
 GN Name=Ewari; Synonym=EWS; EWSH;
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Testis;
 RX MEDLINE=95130099; PubMed=7829090;
 RA Pionasael B., Mattei M.-G., Thomas G., Delattre O.;
 RT "Cloning and chromosome localization of the mouse Ews gene";
 RL Genomics 23:278-281(1994).
 CC -1- FUNCTION: Might function as a repressor (By similarity).
 CC -1- SUBUNIT: Binds RNA, POLR2C, SFI and calmodulin. Interacts with
 CC PTK2B (By similarity).
 CC -1- SUBCELLULAR LOCATION: Nuclear, cytoplasmic and on cell surface.
 CC Relocates from cytoplasm to ribosomes upon PTK2B/PKX activation
 CC (By similarity).
 CC -1- PTM: Phosphorylated; calmodulin-binding inhibits phosphorylation
 CC of Ser-266 (By similarity).
 CC -1- MISCELLANEOUS: Binds calmodulin in the presence, but not in the
 CC absence, of calcium ion (By similarity).
 CC -1- SIMILARITY: Belongs to the RNP TET family.
 CC -1- SIMILARITY: Contains 1 IQ domain.

FT	MOD_RES	309	309		similarity)	Asymmetric dimethylarginine (By similarity)
FT	MOD_RES	314	314		similarity)	Asymmetric dimethylarginine (By similarity)
FT	MOD_RES	317	317		similarity)	Asymmetric dimethylarginine (By similarity)
FT	MOD_RES	321	321		similarity)	Asymmetric dimethylarginine (By similarity)
FT	MOD_RES	454	454		similarity)	Asymmetric dimethylarginine (By similarity)
FT	MOD_RES	463	463		similarity)	Asymmetric dimethylarginine (By similarity)
FT	MOD_RES	470	470		similarity)	Asymmetric dimethylarginine (By similarity)
FT	MOD_RES	489	489		similarity)	Asymmetric dimethylarginine (By similarity)
FT	MOD_RES	493	493		similarity)	Asymmetric dimethylarginine (By similarity)
FT	MOD_RES	499	499		similarity)	Asymmetric dimethylarginine (By similarity)
FT	MOD_RES	502	502		similarity)	Asymmetric dimethylarginine (By similarity)
FT	MOD_RES	505	505		similarity)	Asymmetric dimethylarginine (By similarity)
FT	MOD_RES	562	562		similarity)	Asymmetric dimethylarginine (By similarity)
FT	MOD_RES	564	564		similarity)	Asymmetric dimethylarginine (By similarity)
FT	MOD_RES	571	571		similarity)	Asymmetric dimethylarginine (By similarity)
FT	MOD_RES	574	574		similarity)	Asymmetric dimethylarginine (By similarity)
FT	MOD_RES	580	580		similarity)	Asymmetric dimethylarginine (By similarity)
FT	MOD_RES	588	588		similarity)	Asymmetric dimethylarginine (By similarity)
FT	MOD_RES	591	591		similarity)	Asymmetric dimethylarginine (By similarity)
FT	MOD_RES	595	595		similarity)	Asymmetric dimethylarginine (By similarity)
FT	MOD_RES	602	602		similarity)	Asymmetric dimethylarginine (By similarity)
FT	MOD_RES	606	606		similarity)	Asymmetric dimethylarginine (By similarity)
FT	MOD_RES	614	614		similarity)	Asymmetric dimethylarginine (By similarity)
FT	MOD_RES	632	632		similarity)	Asymmetric dimethylarginine (By similarity)
FT	MOD_RES	635	635		similarity)	Asymmetric dimethylarginine (By similarity)
FT	MOD_RES	266	266		Phosphoserine (By PKC) (By similarity).	
SEQ	SEQUENCE	655 AA;	68418 MW;	50735EDBS4247D69 CRC64;		
QY	Query Match	98.3%;	Score 1888.5;	DB 1;	Length 655;	
Db	Best Local Similarity	98.2%;	Pred. No. 3.3e-101;			
	Matches	332;	Conservative	4;	Mismatches	1;
					Indels	1;
					Gaps	1;
QY	1	GGRGKMSAGERGSGFNKPGGPMDEGPLDLGAPVDPDEDSDNSAIYYGLNDSVTLLDDL A	60			
Db	319	GGRGGLG-AGBRRGGFNRKPGGPMDEGPLDLGFLPIDPDEDSDNSAIYYOGLNDNTTLLDLA	377			
QY	61	DFFPGCGVVKKRKTGGPMTHIYLDKETGKRPKGATSYSDPPAKAAVENFPDGKDFGS	120			
Db	378	DFFPGCGVVKKRKTGGPMTHIYLDKETGKRPKGDATSYSDPPAKAAVENFDKDFGS	437			
QY	121	KLKATSLARKKPENMNMKGILPREGRAMPPLRLRGPGGPGGPGGMGRGGDRGF	180			
Db	438	KLKATSLARKKPENMNMKGMPPRRGRCMPPLRLRGPGGPGGPGGMGRGGDRGF	497			
QY	181	PRGRGRGNPDGGGANVOHRAQMCCNPCCGNGNPFMRRETCDCKAPKEGFLPPPPP	240			

Db	498	PGCPGRSRNPGSGGAVQHRAGMOCENRCCGNQNPAMRTECHQCRAPKEGFLPPFP	557
Qy	241	PGDGRGPGGMRGGRGGGLMDRGFGGFMFRGRCGDGRCGFRGRCMDRGFGGGRG	300
Db	558	PGDGRGRCGPGGMRGGRGGGLMDRGFGGFMFRGRCGDGRCGFRGRCMDRGFGGGRG	611
Qy	301	GGPPPLMOMGMRGRCGRCGRCGRCMDRGRCGRDRPY	338
Db	618	GGPPPLMOMGMRGRCGRCGRCGRCMDRGRCGRDRPY	655
RESULT 7			
Q9D2P0	PRELIMINARY;	PRT;	655 AA.
AC	Q9D2P0;		
DT	01-JUN-2001 (TrEMBLrel. 17, Created)		
DT	01-JUN-2001 (TrEMBLrel. 17, Last sequence update)		
DT	01-MAR-2004 (TrEMBLrel. 26, Last annotation update)		
DE	Mus musculus 18 days pregnant adult female placenta and extra embryonic tissue cDNA, RIKEN full-length enriched library, clone:3830417B11 product:Ewing sarcoma homolog, full insert sequence.		
GN	Name=Ewing; Synonyms=Ewh;		
OS	Mus musculus (Mouse).		
OC	Eukaryota; Metazoa; Chordata; Craniala; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
OX	NCBI_TaxID=10090;		
LN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=C57BL/6J; TISSUE=Placenta and extra embryonic tissue;		
RC	MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;		
RA	Carninci P., Hayashizaki Y.;		
RT	"High-efficiency full-length cDNA cloning.";		
RL	Meth. Enzymol. 303:19-44(1999).		
LN	[2]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=C57BL/6J; TISSUE=Placenta and extra embryonic tissue;		
RC	MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;		
RA	RIKEN FANTOM Consortium;		
RT	"Functional annotation of a full-length mouse cDNA collection.";		
RL	Nature 409:685-699(2001).		
LN	[3]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=C57BL/6J; TISSUE=Placenta and extra embryonic tissue;		
RC	The FANTOM Consortium;		
RA	the RIKEN Genome Exploration Research Group Phase I & II Team;		
RT	"Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs.";		
RL	Nature 420:563-573(2002).		
LN	[4]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=C57BL/6J; TISSUE=Placenta and extra embryonic tissue;		
RC	MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;		
RA	Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M., Komori H., Okazaki Y., Muramatsu M., Hayashizaki Y.;		
RT	"Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes.";		
RL	Genome Res. 10:1617-1630(2000).		
LN	[5]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=C57BL/6J; TISSUE=Placenta and extra embryonic tissue;		
RC	MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;		
RA	Shibata K., Itoh M., Aizawa K., Nagaoke T., Sasaki N., Carninci P., Komori H., Akiyama J., Niehi K., Kiteunai T., Tashiro H., Itoh M., Suni N., Ithil Y., Nakamura S., Hazama M., Nishine T., Harada A., Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K., Fujikawa S., Inoue K., Togawa Y., Izawa M., Ohara E., Wataniki M., Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsumura S., Kawai J., Okazaki Y., Muramatsu M., Inoue Y., Kita A., Hayashizaki Y.;		
RT	"RIKEN integrated sequence analysis (RISA) System-384-format sequencing pipeline with 384 multiplexed cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes.";		
RL	Genome Res. 10:1757-1771(2000).		

RN	[6]
RP	SEQUENCE FROM N.A.
RC	STRAIN=C57BL/6J; TISSUE=Placenta and extra embryonic tissue;
RA	Adachi J., Aizawa K., Akahira S., Akimura T., Arai A., Aono H.,
RA	Araoka T., Bono H., Carninci P., Fukuda S., Fukuishi Y., Furuno M.,
RA	Hanagata T., Hara A., Hayatsu N., Hiramoto K., Hiraoka T., Horii F.,
RA	Iinocenti K., Ishii Y., Itoh M., Itawa M., Kasukawa T., Kato H.,
RA	Kawai J., Kojima Y., Konno H., Kouda M., Koya S., Kurihara C.,
RA	Matsuyama T., Miyazaki A., Nishi K., Nomura K., Numazaki R., Ono M.,
RA	Okazaki Y., Okido T., Owa C., Saito H., Saito R., Sakai C., Sakai K.,
RA	Sano H., Sasai D., Shibata K., Shibata Y., Shinagawa A., Shiraki T.,
RA	Sogabe Y., Suzuki H., Tagami M., Tagawa A., Takahashi F., Tanaka T.,
RA	Tejima Y., Toyota T., Yamamura T., Yasunishi A., Yoshida K., Yoshino M.,
RA	Mutamatsumu M., Hayashizaki Y.,
RL	Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.
DR	EMBL; AK019460; BAB31732.1; -
DR	HSSP; 095218; INOZ.
DR	MGD; MG1:99960; Ewarp1.
DR	GO; GO:0005634; C:nucleus, IDA.
DR	InterPro; IPRO11368; RNA-binding_EWS.
DR	InterPro; IPRO00504; RNA_rec_mot.
DR	InterPro; IPRO01876; Znf_RangBP.
DR	Pfam; PF00076; RMF_1; 1.
DR	Pfam; PF00641; zf-RanBP; 1.
DR	PIRSF; PIRSF002101; RNA-binding_EWS; 1.
DR	SMART; SM00360; RMF; 1.
DR	SMART; SM00547; Znf_RB2; 1.
DR	PROSITE; PS0102; RM; 1.
DR	PROSITE; PS01358; ZF_RANBP2_1; 1.
DR	PROSITE; PS50199; ZF_RANBP2_2; 1.
DR	SEQUENCE 655 AA; 68462 MW; 10C706A92166873 CRC64;
QY	Query Match 98.3%; Score 1888.5; DB 2; Length 655;
QY	Best Local Similarity 98.2%; Pred. No. 3.3e-101;
Matches 332; Conservative 4; Mismatches 1; Indels 1; Gaps 1;	
Db	1 GGRGMSAGRGGEFPNKPGCGPMDEGPLDJGPVDPEDSDNSAIYYQGLNDSTLTDLDA 60
Gy	GGRGGLG-AGERGCFNNKPGCGPMDEGPLDJGLPIDPEDSDNSAIYYQGLNDTNTLDLLA 377
Db	61 DFFPGCVGVKKRKTGTGPMTHIYLDKETGPKKGATSYEDPPAKAAVEPFQSKDPOGS 120
Gy	DFFPGCVGVKKRKTGTGPMTHIYLDKETGPKKGATSYEDPPAKAAVEPFQSKDPOGS 437
Db	121 KLKYSILARKKPKNPSMRGSLPRREGGMPLLRGCPGPGCGPQGMWGMGCGDRSGFP 180
Gy	KLKYSILARKKPKNPSMRGSMMPRRGRMPLLRGCPGPGPQGMWGMGCGDRSGFP 497
Db	181 PRFRGSRGNDSGGSNVQHRAQWQCNPCCGNGCNFMARTECNCQCKAPKDEGFLPPPPP 240
Gy	PRFRGSRGNDSGGSNVQHRAQWQCNPCCGNGCNFMARTECNCQCKAPKDEGFLPPPPP 557
Db	241 PGGRGSGPCGMMGCGGLMDRGPBGCMFRGCRGCRGCRGCRGCRGCRGCRGCRGCRGCRG 300
Gy	PGGRGSGPCGMMGCGGLMDRGPBGCMFRGCRGCRGCRGCRGCRGCRGCRGCRGCRGCRG 617
Db	301 GGPGPLMEQWGRRGRRGCGKMDKEHFOERRDRPY 338
Gy	GGPGPLMEQWGRRGRRGCGKMDKEHFOERRDRPY 655
Db	618 GGPGPLMEQWGRRGRRGCGKMDKEHFOERRDRPY 655
RESULT 8	
Q9CRS5	PRELIMINARY; PRT; 333 AA.
AC	09CRS5.
DT	01-JUN-2001 (TrEMBLrel. 17, Created)
DT	01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT	01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE	Mus musculus t13 days embryo head cDNA, RIKEN full-length enriched library, clone;t330002021 product:Evling sarcoma homolog, full insert sequence. (Fragment).
DE	Name=Ewst1; Synonym=Ewbsh;
GN	Mus musculus (Mouse).
OS	

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OK NCBI_TaxId=10090;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Head;
 RX MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
 RA Carninci P., Hayashizaki Y.;
 RT "High-efficiency full-length cDNA cloning.";
 RL Meth. Enzymol. 303:19-44(1999).
 RN (2)
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Head;
 RX MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;
 RA RIKEN PANTOM Consortium;
 RT "Functional annotation of a full-length mouse cDNA collection.";
 RL Nature 409:685-690(2001).
 RN (3)
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Head;
 RA The PANTOM Consortium;
 RT "The RIKEN Genome Exploration Research Group Phase I & II Team;
 RT "Analysis of the mouse transcriptome based on functional annotation of
 RT 60,770 full-length cDNAs.";
 RL Nature 420:563-573(2002).
 RN (4)
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Head;
 RX MEDLINE=2049374; PubMed=11042159; DOI=10.1101/gr.145100;
 RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
 RA Kono H., Okazaki Y., Muramatsu M., Hayashizaki Y.;
 RT "Normalization and subtraction of cap-trapper-selected cDNAs to
 RT prepare full-length cDNA libraries for rapid discovery of new genes.";
 RL Genome Res. 10:1517-1530(2000).
 RN (5)
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Head;
 RX MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;
 RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,
 RA Kono H., Akiyama J., Nishi K., Kiteunai T., Tashiro H., Itoh M.,
 RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishino T., Harada A.,
 RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
 RA Fujiwaka S., Inoue K., Ogawa K., Izawa M., Ohara E., Watanabe M.,
 RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Maruyama S., Kawai J.,
 RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
 RT "RIKEN integrated sequence analysis (RISA) system-384-format
 RT sequencing pipeline with 384 multicapillary sequencer.";
 RL Genome Res. 10:1757-1771(2000).
 RN (6)
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Head;
 RA Aachti J., Aizawa K., Akahira S., Akimura T., Arai A., Aono H.,
 RA Arawaka T., Bono H., Carninci P., Fukuda S., Fukunishi Y., Furuno M.,
 RA Hanagaki T., Hara A., Hayatsu N., Hiramoto K., Hirakata T., Hori F.,
 RA Imotani K., Ishii Y., Itoh M., Izawa M., Kasukawa T., Kato H.,
 RA Kawai Y., Kojima Y., Kono H., Kouda M., Koya S., Kuribara C.,
 RA Matsuyama T., Miyazaki A., Nishi K., Nomura K., Numazaki R., Ono M.,
 RA Okazaki Y., Okido T., Owa C., Saito H., Saito K., Sakai C., Sakai T.,
 RA Sano H., Sasaki D., Shibata K., Shibata Y., Shingawa A., Shiraki T.,
 RA Sogabe Y., Suzuki H., Tagami M., Tagami M., Takahashi F., Tanaka T.,
 RA Tejima Y., Toya T., Yamamura T., Yasunishi A., Yoshida K., Yoshino M.,
 RA Muramatsu M., Hayashizaki Y.;
 RT "Submitted (Jul-2000) to the EMBL/GenBank/DBJ databases.
 RL EMBL; AK014366; BAB29301.1; -.
 DR HSSP; O95218; INOZ.
 DR MGD; MGI:99960; Ewerl.
 DR GO; GO:0005634; C:nucleus; IDA.
 DR InterPro; IPR000504; RNA_rec_mot.
 DR InterPro; IPR001875; znf_RandDP.
 DR Pfam; PF00076; RRM_1; 1.
 DR Pfam; PF00641; zf_RanBP; 1.
 DR SMART; SM00360; RRM; 1.
 DR SMART; SM00547; znf_RBZ; 1.

DR PROSITE; PS50102; RRM; 1.
 DR PROSITE; PS01358; zf_RanBP2_1; 1.
 DR PROSITE; PS50199; zf_RanBP2_2; 1.
 FT NON_TER
 SQ SEQUENCE 333 AA; 34617 MW; CAEDCA33F2F07 CRC64;
 Query Match 97.5%; Score 1874; DB 2; Length 333;
 Best Local Similarity 98.5%; Pred. No. 1.2e-100;
 Matches 328; Conservative 4; Mismatches 1; Indels 0; Gaps 0;
 QY 6 MSGASRGGFNKGPGPMDEGPDLDGPPVDEDESDNSAIYQGLNDSYTLDDLDFPQ 65
 DB 1 LGSASRGGFNKGPGPMDEGPDLDGPPVDEDESDNSAIYQGLNDSYTLDDLDFPQ 60
 QY 66 CGVVMNKRRTGQPMHIYLDKGTGPKGDATSYSDPPAKAAVEMFGDKDFQSKLXYS 125
 DB 61 CGVVMNKRRTGQPMHIYLDKGTGPKGDATSYSDPPAKAAVEMFGDKDFQSKLXYS 120
 QY 126 LARKKTPMNSMRGGLPPREGGMPPLRLGPGPGPGPGPMGMRGGRGGRGPPRGPR 185
 DB 121 LARKKTPMNSMRGGLPPREGGMPPLRLGPGPGPGPGPMGMRGGRGGRGPPRGPR 180
 QY 186 GSRGNPSSGGVYQHRADGWQCPNPGCQNGNFAMRTCCNQCKAPKDEGLPPPPPGGR 245
 DB 181 GSRGNPSSGGVYQHRADGWQCPNPGCQNGNFAMRTCCNQCKAPKDEGLPPPPPGGR 240
 QY 246 GRRGSGMRGGRGGLMDRGPGMFRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGR 305
 DB 241 GRRGSGMRGGRGGLMDRGPGMFRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGR 300
 QY 306 PLMEQMGRRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGR 338
 DB 301 PLMEQMGRRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGR 333
 RESULT 9
 Q6P3NO PRELIMINARY; PRT; 674 AA.
 ID Q6P3NO
 AC Q6P3NO;
 DT 05-JUL-2004 (TrEMBLrel. 27, Created)
 DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
 DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
 DE Hypothetical protein MGC6258.
 GN Name=MGC6258;
 OS Xenopus tropicalis (Western clawed frog) (Silurana tropicalis).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
 OC Xenopodinae; Xenopus.
 OK NCBI_TaxId=8364;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC TISSUE=Embryo;
 RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
 RA Strausberg R.L., Feilgold B.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shemmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buettner K.H., Scheffer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Schetz T.E.,
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loughell N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
 RA Bosak S.A., McGowan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Morley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Hellon E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green B.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
 RA Krzywicki M.I., Skalka U., Smalls D.E., Scherch A., Schein J.E.,
 RA Jones S.J., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length human
 RT and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).

[2]
 RN SEQUENCE FROM N.A.
 RC TISSUE=Embryo;
 RA Klein S., Gerhard D.S.;
 RL Submitted (DEC-2003) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC063928; AAH63928.1; -
 DR InterPro: IPR006162; Pfam: S.
 DR InterPro: IPR000504; RNA_rec_mol.
 DR InterPro: IPR001876; Znf_RandDP.
 DR Pfam; PF00076; RRM_1; 1.
 DR Pfam; PF00641; zf-RanBP; 1.
 DR SMART; SMO0360; RRM; 1.
 DR SMART; SMO0547; Znf_RBZ; 1.
 DR PROSITE; PS00012; PHOSPHOPANTHETHEINE; UNKNOWN_1.
 DR PROSITE; PS01358; ZF_RANBP2_1; 1.
 DR PROSITE; PS01358; ZF_RANBP2_2; 1.
 DR Hypothetical protein.
 KW Hypothetical protein.
 SQ SEQUENCE 674 AA; 71082 MW; C4B34C7B95449B0D CRC64;
 Query Match 69.5%; Score 1336; DB 2; Length 674;
 Best Local Similarity 72.0%; Pred. No. 2,4e-69;
 Matches 250; Conservative 37; Mismatches 34; Indels 26; Gaps 11;
 QY 1 GGRGNGSAGEGFGNKGPMDEGPDLDGP-----PVPDESDNSAIYVQGLNDV 54
 DB 345 GSRGNGSGGDRAGFGSLDDGPDLDGPMPPLPLPLDLDELSSTIYVQGLNDV 404
 QY 55 TLDDLADFPKQCGVVMKRTGQPMHILYDKETGPKGDATVSYEDPPTAKAAVWFQD 114
 DB 405 TVBEIVDFHCGDVINKRTGSELVNLPMDEKTKGKSGTYSFEDPSAKTAIELCDG 464
 QY 115 KDFQSKLVSLARKKPPMNSMRGGLPREGGMPPLRLGCGPGGPGGPMGRMGGRG 174
 DB 465 KDLNGKVKVSLARKKSLGSMRGSLRLDNRGQPPPLR-----GGPMRLGGRG 515
 QY 175 DRGGPPRGRGSRGNPSGGVNVHRAGDMQCPNPGCGNQNFAMRTCCNQCKAPKPDG 234
 DB 516 ERGGFMRGPRGRGSPV-SGNVQHRAGDMQCPNPGCGNQNFAMRTCCNQCKAPKPDG-- 572
 QY 235 P-PPPPPGDGRGPGMRGGRGGLMDRGPGGMRGGRGGRGGRGGRGGRGGRG 293
 DB 573 PPPPPPGGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGR 627
 QY 294 GGRGGRGPGPPPLMEOMGRRG--RGPGKMDKGEHRRDRPY 338
 DB 628 GGRGGRGPGPPPLMEOLGGRIGGGRGGRGGRGGRGGRGGRGGRGGRGGRG 674
 RESULT 10
 ID Q6DUQ4 PRELIMINARY; PRT; 673 AA.
 AC Q6DUQ4;
 DT 25-OCT-2004 (Tremblrel. 28, Created)
 DT 25-OCT-2004 (Tremblrel. 28, Last sequence update)
 DT 25-OCT-2004 (Tremblrel. 28, Last annotation update)
 DE Hypothetical protein MGC76258.
 GN Name:MGC76258;
 OS Xenopus tropicalis (western clawed frog) (Silurana tropicalis).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipridae; Pipidae;
 OC Xenopodinae; Xenopus.
 OC NCB1_TaxID=8364;
 RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
 RP TISSUE=Embryo;
 RA Strausberg R.L., Pelting E.A., Grouse L.H., Derge J.G., Schuler G.D.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Hootch R.F., Jordan H., Moore T., Max S.I., Wang J., Heile F.,
 RA Diatchenko L., Marisla K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Schetz T.E.,

RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loguettano N.A., Peters G.U., Abramson R.D., Mullaly S.J.,
 RA Bosak S.A., McSwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Morley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Paley J., Helton E., Ketterman M., Madan A., Rodriguez S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
 RA Krzyzanski M.I., Skalka U., Smalish D.E., Scherch A., Schein J.E.,
 RA Jones S.J., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length human
 RT and mouse cDNA sequences";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Embryo;
 RA Klein S., Gerhard D.S.;
 RL Submitted (JUN-2004) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC075120; AAH75120.1; -
 DR GO; GO:0005634; Cnucleus; IEA.
 DR GO; GO:0003676; F-nucleic acid binding; IEA.
 DR GO; GO:0008270; F-zinc ion binding; IEA.
 DR InterPro: IPR006162; Pfam: S.
 DR InterPro: IPR001368; RNA-binding_EWS.
 DR InterPro: IPR000504; RNA_rec_mol.
 DR InterPro: IPR001876; Znf_RandDP.
 DR Pfam; PF00076; RRM_1; 1.
 DR Pfam; PF00641; zf-RanBP; 1.
 DR PIRSF; PIRSF002101; RNA-binding_EWS; 1.
 DR SMART; SMO0360; RRM; 1.
 DR SMART; SMO0547; Znf_RBZ; 1.
 DR PROSITE; PS00012; PHOSPHOPANTHETHEINE; UNKNOWN_1.
 DR PROSITE; PS01358; ZF_RANBP2_1; 1.
 DR PROSITE; PS01358; ZF_RANBP2_2; 1.
 DR Hypothetical protein.
 KW Hypothetical protein.
 SQ SEQUENCE 673 AA; 70995 MW; BD8A817F25991A74 CRC64;
 Query Match 68.8%; Score 1321.5; DB 2; Length 673;
 Best Local Similarity 71.8%; Pred. No. 1.6e-68;
 Matches 249; Conservative 37; Mismatches 34; Indels 27; Gaps 12;
 QY 1 GGRGNGSAGEGFGNKGPMDEGPDLDGP-----PVPDESDNSAIYVQGLNDV 54
 DB 345 GSRGNGSGGDRAGFGSLDDGPDLDGPMPPLPLPLDLDELSSTIYVQGLNDV 403
 QY 55 TLDDLADFPKQCGVVMKRTGQPMHILYDKETGPKGDATVSYEDPPTAKAAVWFQD 114
 DB 404 TVBEIVDFHCGDVINKRTGSELVNLPMDEKTKGKSGTYSFEDPSAKTAIELCDG 463
 QY 115 KDFQSKLVSLARKKPPMNSMRGGLPREGGMPPLRLGCGPGGPGGPMGRMGGRG 174
 DB 464 KDLNGKVKVSLARKKSLGSMRGSLRLDNRGQPPPLR-----GGPMRLGGRG 514
 QY 175 DRGGPPRGRGSRGNPSGGVNVHRAGDMQCPNPGCGNQNFAMRTCCNQCKAPKPDG 234
 DB 516 ERGGFMRGPRGRGSPV-SGNVQHRAGDMQCPNPGCGNQNFAMRTCCNQCKAPKPDG-- 571
 QY 235 P-PPPPPGDGRGPGMRGGRGGLMDRGPGGMRGGRGGRGGRGGRGGRGGRGGRG 293
 DB 572 PPPPPPGGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGR 626
 QY 294 GGRGGRGPGPPPLMEOMGRRG--RGPGKMDKGEHRRDRPY 338
 DB 627 GGRGGRGPGPPPLMEOLGGRIGGGRGGRGGRGGRGGRGGRGGRGGRGGRG 673
 RESULT 11
 ID Q6NIX1 PRELIMINARY; PRT; 624 AA.
 AC Q6NIX1;
 DT 05-JUL-2004 (Tremblrel. 27, Created)

DT 05-JUL-2004 (TrEMBLrel. 27, last sequence update)
 DT 05-JUL-2004 (TrEMBLrel. 27, last annotation update)
 DE Wu:fc04c01 protein (Fragment).
 GN ORFNames=wu:fc04c01;
 OS Brachydanio rerio (Zebrafish) (Danio rerio).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
 OC Cyprinidae; Danio.
 NC NCBITaxID=7955;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Embryo;
 RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., McSwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hultyn S.W.,
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahy J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
 RA Krzywiński M.I., Skalska U., Smalls D.E., Schnerch A., Schein J.E.,
 RA Jones S.J., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length human
 and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Embryo;
 RX Submitted (APR-2004) to the EMBL/GenBank/DBJ databases.
 RA Strausberg R.;
 RA EMBL; BC068397; AAH68397.1; -; wu:fc04c01.
 RA ZFIN; ZDB-GENE-030131-2317; wu:fc04c01.
 DR InterPro; IPR005054; RNA_rec.mot.
 DR InterPro; IPR005829; Sug_transporter.
 DR InterPro; IPR001876; Znf_RanBP.
 DR Pfam; PF00076; RRM_1; 1.
 DR Pfam; PF00641; zf-RanBP; 1.
 DR SMART; SM00360; RRM; 1.
 DR SMART; SM00547; ZNF_RBZ; 1.
 DR PROSITE; PS00102; RRM; 1.
 DR PROSITE; PS00216; SUGAR_TRANSPORT_1; UNKNOWN_1.
 DR PROSITE; PS01358; ZF_RANBP2_1; 1.
 DR PROSITE; PS01359; ZF_RANBP2_2; 1.
 FT NON_TER
 SQ SEQUENCE 624 AA; 64731 MW; D0113DF02861F221 CRC64;

Query Match 61.8%; Score 1187.5; DB 2; Length 624;
 Best local Similarity 67.9%; Pred. No. 8e-61;
 Matches 235; Conservative 33; Mismatches 42; Indels 37; Gaps 15;

1 GG--RGGMSGAGRGFGNFKGFGPMDEGPDIDGPPVDPDSDNSAIVYQGLNDVTLDD 58
 308 GGSRRGSGAGJAGDRGFGSKPG-----DGMGAPERQD--DSHNSIYITGLTENTLLE 359
 59 LADPFQCGVVKNNKRTGGMIMHLYDKNGKPKGATVYEPDPAKAVEMPDGDPQ 118
 360 VADPFQSGIIRINKRTGLPAVNIYTDKDGKPGDATTLSYEPPSKAAVEMFDGDPQ 419
 119 GSKLTKSLARKKPPNNMVRGGLPPREGRGMPPLRGPGGPGGPGGPMGMG--GRGGDRG 177
 420 GKTLTKSLARKKPPNNMVRGGLPPREGRGMPPLRGPGGPGGPGGPMGMG--GRGGDRG 469

178 GPPPR-GPRG-SHGNSGSGGVQHRAGDMQCPNPGCGNQPFAVTECNQCKAPRPGFLP 235
 470 GFMPRGSGPRMGGRGPT--GGNMQORAGDMQCPNAGCGNQFAVMECNQCKAPRPGFGP 528

QY 236 PPPPPGDRGGRGGRGGRGGLMDRGPGGM--FRGGRGDRGFRGGRGMRGGRG 293
 DB 529 PPF-PPGDRGGRGGRGGRGGRG--MDRGFRGPGGFRGGRVNDGRR--GGMGRGGRG 584
 QY 294 GGRGPGGPPPLMEQMGRRGRGGRGPGKMD--KEHRQERRRPPY 338
 DB 585 G---RGRGPF---MDMGRRGRGGRGPPGKMDKDHQDRRRRPPY 624

RESULT 12
 ID Q7SZR6 PRELIMINARY; PRT; 623 AA.
 AC Q7SZR6;
 DT 01-OCT-2003 (TrEMBLrel. 25, Created)
 DT 01-OCT-2003 (TrEMBLrel. 25, last sequence update)
 DT 01-MAR-2004 (TrEMBLrel. 26, last annotation update)
 DE Wu:fc04c01 protein (Fragment).
 GN ORFNames=wu:fc04c01;
 OS Brachydanio rerio (Zebrafish) (Danio rerio).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
 OC Cyprinidae; Danio.
 NC NCBITaxID=7955;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Embryo;
 RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., McSwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hultyn S.W.,
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahy J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
 RA Krzywiński M.I., Skalska U., Smalls D.E., Schnerch A., Schein J.E.,
 RA Jones S.J., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length human
 and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Embryo;
 RX Submitted (AUG-2003) to the EMBL/GenBank/DBJ databases.
 RA Strausberg R.;
 RA EMBL; BC056281; AAH56281.1; -;
 RA HSSP; Q9Y589; 1P27.
 DR ZFIN; ZDB-GENE-030131-2317; wu:fc04c01.
 DR InterPro; IPR005054; RNA_rec.mot.
 DR InterPro; IPR005829; Sug_transporter.
 DR InterPro; IPR001876; Znf_RanBP.
 DR Pfam; PF00076; RRM_1; 1.
 DR Pfam; PF00641; zf-RanBP; 1.
 DR SMART; SM00360; RRM; 1.
 DR SMART; SM00547; ZNF_RBZ; 1.
 DR PROSITE; PS00102; RRM; 1.
 DR PROSITE; PS00216; SUGAR_TRANSPORT_1; UNKNOWN_1.
 DR PROSITE; PS01358; ZF_RANBP2_1; 1.
 DR PROSITE; PS01359; ZF_RANBP2_2; 1.
 FT NON_TER
 SQ SEQUENCE 623 AA; 64517 MW; BE47E2ED7FEB9F60 CRC64;

Query Match 60.7%; Score 1166.5; DB 2; Length 623;
 Best local Similarity 67.8%; Pred. No. 1.3e-59;
 Matches 232; Conservative 31; Mismatches 42; Indels 37; Gaps 15;


```

QY 1 GG--RGMSAGRGGFNKRPGPMDEGPDLDLGPVDPEDSDNSAIYVQGLNDSTVLLD 58
DB 311 GGMSRGMGAGRGGRGSKPG-----DSBMGAPEEPQD--DSBNSITYITGLTENATLER 362
QY 59 LADFFKQCGVKNKRTGQPMHIYLDKETGPKGDATSYEDPPTAKAVWFPGDKDFQ 118
DB 363 VADFFHSGSIRINRKTGLPAVNIYTDKTKGPKGDATLSYEPSPAKAVWFPGDKDFQ 422
QY 119 GSKLTKSLARKKPPMNSMRGGLPPRRGRGMPPLRGPGGPGGPGGPMWG--GRGDRG 177
DB 423 GKKLTKSLARKKPPMNSMRGGLPPRRGRGMPPLRGPGGPGGPGGPMWG--GRGDRG 472
QY 178 GPRPR--GPRG--SRGNSGCGNVQHRAGDMQCPNPGCGNQNFAWRTGNCOKAKPFGFLP 235
DB 473 GFMPRGGRGPRGMRGGRGPT--GGNMQRAGDMQCPNAGCGNQNFAWRTGNCOKAKPFGFLP 531
QY 236 PPPPPGDRGRGGRGPMRGGRGGLMDRGPGGM--FRGGRGDRGGRGGRGGRGGRG 293
DB 532 PPR--PRGDRGRGGRGPMRGGRG--MDRGPGGPGGFRGGRGGRGGRGGRGGRG 587
QY 294 GGRGGRGPGGPPGLMEQMGRRGGRGGRGGRGMD--KGEHROERR 334
DB 588 G---RGRGPP---MDMGRGRGRGMPGPKMDKGDHRODR 623

RESULT 13
Q6NMA7 PRELIMINARY; PRT; 575 AA.
AC 06NMA7;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE Wu:fb75g09 protein (Fragment).
GN Name:fb75g09 protein (Fragment).
OS Brachydanio rerio (Zebrafish) (Danio rerio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
NCBI_TaxId=7955;
OX [1]
RN SEQUENCE FROM N.A.
RP TISSUE=Embryo;
RC MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.P., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Haile F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Ueda T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalek U., Smalins D.E., Schnerch A., Schein J.B.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences."
RT Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Embryo;
RA Strausberg R.;
RA Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC067661; AAH67661.1; -.
DR ZFIN; ZDB-GENE-030131-1600; fuzl.
DR InterPro; IPR000504; RNA rec mot.
DR InterPro; IPR001876; znf_RandDP.

```

```

DR Pfam; PF00076; RRM_1; 1.
DR Pfam; PF00641; zf_RanBP; 1.
DR SMART; SM00360; RRM; 1.
DR SMART; SM00547; znf_RBZ; 1.
DR PROSITE; PS50102; RRM; 1.
DR PROSITE; PS01358; zf_RanBP2_1; 1.
DR PROSITE; PS50199; zf_RanBP2_2; 1.
DR NON_TER 1
SQ SEQUENCE 575 AA; 60365 MM; A1650072377A7629 CRC64;

Query Match 55.8%; Score 1073; DB 2; Length 575;
Best Local Similarity 61.9%; Pred. No. 2,9e-54;
Matches 211; Conservative 33; Mismatches 59; Indels 38; Gaps 10;

QY 2 GRGMSAGRGGFNKRPGPMDEGPDLDLGPVDPEDSDNSAIYVQGLNDSTVLLD 61
DB 269 GRGMSAGRGGFNKRPGPMDEGPDLDLGPVDPEDSDNSAIYVQGLNDSTVLLD 326
QY 62 FFKQCGVKNKRTGQPMHIYLDKETGPKGDATSYEDPPTAKAVWFPGDKDFQ 121
DB 327 FFKHTAIRINRRLNPPALNITYDKSGKPKGDATLSYEPAPAKAVWFPGDKDFQ 386
QY 122 LKYSLARKKPPMNSMRGGLPPRRGRGMPPLRGPGGPGGPGMGRGGRGGRG 181
DB 387 LKYSLARKKPPMNSMRGGLPPRRGRGMPPLRGPGGPGGPGMGRGGRGGRG 430
QY 182 R--GPRGSRGNPSC--GGNVQHRAGDMQCPNPGCGNQNFAWRTGNCOKAKPFGFLP 237
DB 431 RGPRGGMGNNGPQGVQVQKRAWDCEPNAGCGNQNFAWRTGNCOKAKPFGFLP 490
QY 238 PPPPGDRGRGGRGPMRGGRGGLMDRGPGGM--FRGGRGDRGGRGGRGGRG 297
DB 491 F--SPGRGR--GGRGGRG--MDRGPGGM--RGWGDGRGGRGGRGGRGGRG 543
QY 298 GGRGGRGPGGPPGLMEQMGRRGGRGGRGGRGMD--KGEHROERR 338
DB 544 GGP-----PMDGRGRGMPGPKMDKGDHRODR 575

RESULT 14
Q803E3 PRELIMINARY; PRT; 578 AA.
AC Q803E3;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Zgc:55864 protein.
GN Name:fb75g09 protein (Fragment).
OS Brachydanio rerio (Zebrafish) (Danio rerio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
NCBI_TaxId=7955;
OX [1]
RN SEQUENCE FROM N.A.
RP STRAIN=AB; TISSUE=Whole body;
RC MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.P., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Haile F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Ueda T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalek U., Smalins D.E., Schnerch A., Schein J.B.,

```


GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - nucleic search, using frame_plus.p2n model

Run on: February 20, 2005, 16:33:24 ; Search time 3615.31 Seconds
(without alignments)
4530.143 Million cell updates/sec

Title: US-10-791-017a-2_COPY_319_656
Perfect score: 1922
Sequence: 1 GGRGCMGSGRGRGFRKPG.....GGRGKDKGHRGRDRRY 338

Scoring table:
BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 4708233 seqs, 24227607955 residues
Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0
Maximum DB seq length: 200000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODE=frame+ .p2n.model -DEV=xlp
-Q/cgml_1/USPTO.spool_p/US10791017/runat.17022005.125807.22074/app.query.fasta_1.1358
-DB=GenEmbl -QFMT=fastap -SUFFIX=p2n.rge -MINMATCH=0.1 -LOOPEL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blomsu62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=prc -NOR=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=200000000
-USER=US10791017_@CGML_1.1.7357_@runat.17022005.125807.22074 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLCK=100 -LONGLOG
-DEV_TIMEOUT=120 -MARN_TIMEOUT=30 -THREDS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl.1.*
1: gb_ba.*
2: gb_hcg.*
3: gb_in.*
4: gb_com.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pl.*
9: gb_pr.*
10: gb_ro.*
11: gb_ste.*
12: gb_sy.*
13: gb_un.*
14: gb_vl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1922	100.0	1971	12 BT007796	BT007796 Synthetic
2	1922	100.0	1988	6 AX714390	AX714390 Sequence
3	1922	100.0	1988	9 AK056681	AK056681 Homo sapi
4	1922	100.0	2182	9 BC004817	BC004817 Homo sapi

5	1922	100.0	2371	6 A36460	A36460 Sequence 1
6	1922	100.0	2371	6 AR080100	AR080100 Sequence
7	1922	100.0	2390	6 CQ867364	CQ867364 Sequence
8	1922	100.0	2390	6 AX411125	AX411125 Sequence
9	1922	100.0	2390	6 HSEMS	X66899 H.sapiens E
10	1909	99.3	2026	6 CQ850483	CQ850483 Sequence
11	1909	99.3	2026	6 AK127624	AK127624 Homo sapi
12	1909	99.3	2326	6 CQ730352	CQ730352 Sequence
13	1907.5	99.2	2164	6 BC072442	BC072442 Homo sapi
14	1907.5	99.2	2189	9 AK056309	AK056309 Homo sapi
15	1907.5	99.2	2189	9 CR456490	CR456490 Homo sapi
16	1907.5	99.2	2364	9 BC011048	BC011048 Homo sapi
17	1903	99.0	2189	10 BC068226	BC068226 Mus muscu
18	1888.5	98.3	2188	6 AX305537	AX305537 Sequence
19	1888.5	98.3	2188	10 MMEMS	X79233 M.musculus
20	1888.5	98.3	2396	5 BC083960	BC083960 Xenopus 1
21	1878	97.7	7293	9 HSM808920	BX648769 Homo sapi
22	1816	94.5	2473	9 AK026270	AK026270 Homo sapi
23	1678	87.3	221444	2 AC113313	AC113313 Mus muscu
24	1678	87.3	247757	2 AC121282	AC121282 Mus muscu
25	1676	87.2	2440	5 AJ719366	AJ719366 Gallus ga
26	1661	86.4	132906	2 AC109802	AC109802 Canis fam
27	1661	86.4	163104	2 AC110672	AC110672 Canis fam
28	1632	84.9	193559	2 AC137500	AC137500 Homo sapi
29	1619	84.2	155815	2 AC018774	AC018774 Homo sapi
30	1619	84.2	180718	9 AL596087	AL596087 Human DNA
31	1619	84.2	182501	2 AC011221	AC011221 Homo sapi
32	1541	80.2	1783	6 CQ721057	CQ721057 Sequence
33	1443.5	75.1	226400	2 AC128482	AC128482 Rattus no
34	1443.5	75.1	235241	2 AC098823	AC098823 Rattus no
35	1443.5	75.1	263925	2 AC106522	AC106522 Rattus no
36	1406.5	73.2	2327	9 AK096787	AK096787 Homo sapi
37	1385	72.1	1462	6 CQ434162	CQ434162 Sequence
38	1336	69.5	2665	5 BC063928	BC063928 Xenopus t
39	1321.5	68.8	2683	5 BC075120	BC075120 Xenopus t
40	1187.5	61.8	2449	5 BC068397	BC068397 Dario rer
41	1181.5	61.5	2448	5 BC056281	BC056281 Dario rer
42	1165.5	60.6	270337	2 AC099212	AC099212 Rattus no
43	1105	57.5	221924	2 AC105889	AC105889 Rattus no
44	1105	57.5	225105	2 AC118907	AC118907 Rattus no
45	1073	55.8	2093	5 BC044518	BC044518 Dario rer

ALIGNMENTS

RESULT 1
BT007796 1971 bp mRNA linear SYN 13-MAY-2003
LOCUS Synthetic construct Homo sapiens Ewing sarcoma breakpoint region 1
DEFINITION mRNA, partial cds.
ACCESSION BT007796 GI:30584430
VERSION BT007796.1
KEYWORDS FLI CDNA.
SOURCE synthetic construct
ORGANISM synthetic construct
REFERENCE 1 (bases 1 to 1971)
AUTHORS Kainline,N., Chen,X., Rolfs,A., Halleck,A., Hines,L., Eisenstein,S., Koundinya,M., Raphael,J., Moreira,D., Kelley,T., Labaer,J., Lin,Y., Phelan,M. and Farmer,A.
TITLE Cloning of human full-length CDSs in BD Creator(TM) System Donor vector
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1971)
AUTHORS Kainline,N., Chen,X., Rolfs,A., Halleck,A., Hines,L., Eisenstein,S., Koundinya,M., Raphael,J., Moreira,D., Kelley,T., Labaer,J., Lin,Y., Phelan,M. and Farmer,A.
TITLE Direct Submission
JOURNAL Submitted (13-MAY-2003) BD Biosciences Clontech, 1020 East Meadow Circle, Palo Alto, CA 94303, USA
COMMENT This CDS clone is a part of a collection of human full length expression clones generated by BD Biosciences Clontech and the Harvard Institute of Proteomics. Each CDS has been cloned in two

forms: wild and without stop-codon (to allow fusion with C-terminal tag). The CDS has been directionally cloned using BP In-Fusion (TM) cloning system between the SalI and HindIII sites of the pDNR-DUAL vector. Additional sequences in the clone: 'ATC' after SalI site and before 'ATC' to provide Kozak consensus sequence; 'GG' after last codon and before HindIII site to maintain reading frame. Last clone distribution: <http://bioinfo.clonetechn.com/orfclone>.

FEATURES

Source

```

source
1. .1971
/organism="synthetic construct"
/mol_type="mRNA"
/db_xref="taxon:32630"
/clone="GH00297L1.0"
/clone_1b="BD Creator (TM) CDS Library derived from MGC
collection"
/lab_host="DH5alpha Ti resistant"
/note="Vector: pDNR-Dual"
1. .>1971

```

CDS

[illegible]

ORIGIN

Alignment Scores:

Pred. No.:	3,1e-7	length:	197
Score:	1922.00	Matches:	338
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	12	Gaps:	0

US-10-791-017A-2_COPY_319_656 (1-338) x BT007796 (1-1971

QY	I	GIYGIYVRGSLVGIYMECGLYSERIALAGIYGIUATRGSLVGIYPHAAHLYPPIRGILGIY	20
Db	955	GGAGGAGCGCGGTGAMATGGGCACCGCTGGAGAGCGAGGGCTTCAATTAACCTGGATGGA	101
QY	21	ProMetAspGIUGIYProAspPLeuAspLeuGIYProProValAspProAspGIUAspSer	40
Db	1015	CCCATGGATGAAGGACCAAGATCTTGATCTTAGGCCCACTGTGATGCCAGATGAAGACTCT	107
QY	41	AspAsnSerAlaIleTYValGInGIYLeuAsnAspSerValThrLeuAspAspLeuAla	60
Db	1075	GACAACAGTGCATTTATGTATACAGAGATTAAATGAACAGTGACTCTAGATGATCTGGCA	113
QY	61	AspPhePheIysGInCYRGILYValValValYMetAsnIlyAsnArgThyGIYGIYProMetIle	80
Db	1135	GACCTCTTTAAGCAGTGTGGGGTGTGAATGAATGAACAAAGAACTGGCCAACTCCATGATC	119
QY	81	HisIleTYrLeuAspIlySGIYThrGIYLYSPProIysGIYAspAlaThsValSerTYrGIU	100
Db	1195	CACATCTTACCTGGACAAAGAAACGAAAGCCCAAAAGCGATGCCACAGTGTCTTATGA	125
QY	101	AspProProThralaIyValaaIaValGIUTripPheAspGIYLYAspPheGInGIYSer	120
Db	1255	GACCAACCCACCTGCCAAAGGCTGGCCGTGGATGTGTTGATGGGAAAGATTTTCAAGGAGC	133
QY	121	LYaLeuIyValSerLeuAlaHGIYLYbIySPProMetAsnSerMechGIYGIYLeu	140

Db	1315	AAACTTAAAGTCTCCCTTGCCTCGGAAGAGCCCTCCATATGAACAGTATGCGGGGTGCTGC	1374
Qy	141	ProProArgrglucllyArgrglymeProProProleuArgrgllygllyProgllyProglly	160
Db	1375	CCACCCTGGTAGGGACAGAGGCAATGCCACCACTCCGTGAGGTCCAGAGGCCCAAGGA	1434
Qy	161	GlyProogllygllyPrometgllyArmetgllygllyArgrgllygllyArbArgrgllygllyPhePro	180
Db	1435	GGTCTCTGGGGGACCCCATGGGTCCCATATGGAGAGCGCGTGGAGAGATAGAGGAGGCTTCCT	1494
Qy	181	ProArgrgllyProArgrgllySerArgrgllyArbProSerlgllygllyArbValGlnHiaArgr	200
Db	1495	CCAAGAGAGACCCCGGGGGTTCGCCAGGGGAAACCCCTCTTGAGAGAGGAAACGTCCAGACCGA	1554
Qy	201	AlAGlyArbPrrpogllyncysProArbProProgllyCyGlylyArbnglnArbArbPheAlArbArgrThr	220
Db	1555	GCTGGAGACCTGGACAGTGTCCCAATCCGGATTGGGAAACAGAACTTCGCTGGAACA	1614
Qy	221	GlyCyArbArnglnCyBlyArbAlArbProlYArProgllygllyPheLeuProProProPheProPro	240
Db	1615	GAGTGCACACAGTGTAAAGGCCCAAGCCTGMAAGCTTCCTCCGCCACCTTTCCGCC	1674
Qy	241	ProogllygllyArbArgrgllyArgrgllygllyProgllygllyArbArgrgllygllyArgrgllygllyLeu	260
Db	1675	CCGGGTGTGTATCGTGGCAAGGTGGCCCTGTGTGGCATGCGGGGAGGAAAGGTGGGCTTC	1734
Qy	261	MetArbArgrgllygllyProProgllygllyMetPheArgrgllygllyArgrgllygllyArbArgrgllyglly	280
Db	1735	ATGATGTGTGTGTGTCCCGGTGAAATTTCAAGGTGGCCGTGTGGAGACAGAGGTGGC	1794
Qy	281	PheArgrgllygllyArbArgrgllyMetArbArgrgllygllyPhegllygllygllyArbArgrgllygllyPro	300
Db	1795	TTCCGTGTGTGGCCGGGCAATGACCCAGGTGGCTTTGTGTGGAGAAACAGAGTGGCCCT	1854
Qy	301	GlygllyProProProgllyProleuMetGlnGlnMetgllygllyArbArgrgllygllyArgrgllyglly	320
Db	1855	GGGGGGCCCCCTGGACCTTTGATGGAAACAATGGGAGAAAGAGAGAGGACGTGGAGGA	1914
Qy	321	ProgllylyArbMetArblygllygllyGlnHiaArgrgllygllyArbArgrgllygllyProTyr	338
Db	1915	CTTGAAAAAATGATTAAGCGAGCACGTCAGAGGCGAGAGATCGGCGCTTAC	1968
RESULT 2			
AX714390		1988 bp	DNA linear PAT 15-APR-2003
LOCUS	AX714390		
DEFINITION	Sequence 1074 from Patent EP1293569.		
ACCESSION	AX714390		
VERSION	AX714390.1		
KEYWORDS	GI:29889342		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniota; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
AUTHORS	1 Isega,T., Sugiyama,T., Otsuki,T., Wakamatsu,A., Sato,H., Iehi,S., Yamamoto,J.T., Isono,Y., Hio,Y., Otsuka,K., Nagai,K., Irie,R., Tamechika,T., Seki,N., Yoshikawa,T., Otsuka,M., Nagahara,K. and Masuhara,Y.		
TITLE	Full-length cDNA		
JOURNAL	Patent: JP 1293569-A 1074 19-MAR-2003; Helix Research Institute (JP) ; Research Association for Biotechnology (JP)		
FEATURES	Location/Qualifiers		
source	1..1988 /organism="Homo sapiens" /mol_type="unassigned DNA" /db_xref="taxon:9606"		
ORIGIN			
Alignment Scores:	3,12e-47	Length:	1988
Pred. No.:	1922.00	Matches:	338

Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatch:	0
Query Match:	100.00%	Indels:	0
Ds:	6	Gaps:	0

KEYWORDS	ORGANISM	TITLE	JOURNAL	COMMENT	FEATURES	ORIGIN
Homo sapiens (human).	Homo sapiens					
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.						
1 (bases 1 to 2371)						
Autlaas,A., Delattre,O., Desmaza,C., Melot,T., Peter,M., Plougastel,B., Thomas,G. and Zucman,J.						
NUCLEOTIC ACID CORRESPONDING TO A GENE OF CHROMOSOME 22 INVOLVED IN RECURRENT CHROMOSOMAL TRANSLOCATIONS ASSOCIATED WITH THE DEVELOPMENT OF CANCEROUS TUMORS						
Patent: WO 9323549-A 1 25-NOV-1993;						
CENTRE NAT RECH SCIENT (FR)						
Other publication JP 8500964T 960206.						
Location/Qualifiers						
1..2371						
/organism="Homo sapiens"						
/mol_type="unassigned DNA"						
/db_xref="taxon:9606"						
/dev_stage="FOETUS"						
Alignment Scores:						
Pred. No.:	3.5e-47	Length:	2371			
Score:	1922.00	Matches:	338			
Percent Similarity:	100.00%	Conservative:	0			
Best Local Similarity:	100.00%	Mismatches:	0			
Query Match:	100.00%	Indels:	0			
DB:	6	Gaps:	0			
US-10-791-017A-2_COPY_319_656 (1-338) x A26460 (1-2371)						
QY	1	GLYGLYARGGGLYGLYMETGLYSERLAAGLYLUAAGGLYGLYPHEASLTYPRGGIYGLY	20			
DB	979	GGAGGACCGCGGTGGAGATGGGACCGCTGGAGGCGAGGTGGCTTCAATTAACCTGGTGG	1038			
QY	21	PROMETASPGIUGLIYPROASPLEUASPLEUGIYPROPOVALAASPPOASPGIUNASPSER	40			
DB	1039	CCCATGATGAGAGGACCAAGATCTTATCTTATGGCCCTCTGTAGATCCAGATGAGACCTCT	1098			
QY	41	ASPAASERLAALIELTYVALGINGLYLEUASNASPSERVALTHILEUASASPLEUALA	60			
DB	1099	GACAAACAGTGCATATTATGTACCAAGATTAAATGACAGTGTGACTCTAGATGATCTGGCA	1158			
QY	61	ASPHENHELYSGINCYGLYVALYLYSHELANLYBAAGTHGLYGLINPROMETILE	80			
DB	1159	GACTTCCTTAAAGCAATGTGGGGTGTAAATGAAACAAAGAACTGGGCAACCAATGATTC	1218			
QY	81	HSLIELTYLLEUASPLEYGLIUTHRGILYSPROLYSGIYAAPALATHVALSERTYGLU	100			
DB	1219	CACATCTACCTGGACAAAGAAACAGAAACCCCAAGGCGATGCCACAGTGTCTTATATA	1278			
QY	101	ASPPROPTOTHALALYALALALAVAGLUTRPPHEASPGIYLYASPPHEGLINGLYSER	120			
DB	1279	GACCCACCACTGCGCAAGGCTGCGCTGGAAATGTTTGAATGGAAAGATTTTCAAGGAGGC	1338			
QY	121	LYSLLEUYSVALSERLEUALLARYGLYSLYSPROMETANSENSEMETARYGLIYGLYEU	140			
DB	1339	AAACTTAAAGTCTCCCTTGCTCTGGAAAGACCTTCAATGAACAGTATCGGGGGTGTG	1398			
QY	141	PROPTARGLUGLIYARGGLYMETPROPTARGLYGLYGLYGLYGLYGLYGLYGLYGLY	160			
DB	1399	CAACCCCGTGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG	1458			
QY	161	GLYPROGLIYGLYPROMETGLIYARGMETGLIYLAARGGLYGLYGLYGLYGLYGLYGLY	180			
DB	1459	GCTCTGGGGGAGCCATGAGGTGTGATGGGAGGCGCTGGAGGAGATGAGGAGGCTTCCCT	1518			
QY	181	PROAARGIYPROAARGIYSERARGIYANPROSERGIYGLIYGLYGLYGLYGLYGLYGLY	200			
DB	1519	CCAAAGAGACCCCGGGGTTCCCGAGAGGAACCCCTTGGAGGAGGAGAAAGTCCAGACCGA	1578			

QY	201	AlaGlyAspTrpGlnGlyProAsnProGlyCyGAGLYAsnGlnAsnPheAlaTPAqThr	220
Db	1579	GCTGGAGACTGGACGTGTCCCAATCCGGGTTGTGGAAACAGAACTTCGCTGGAGAAC	1638
QY	221	GluCyAsnGlnGlyGlyAsnAlaProGlyPProGlnGlyPheLeuProProPheProPro	240
Db	1639	GAGTGCACACAGTGTAAAGGCCCAAGCCGTAAGGCTTCCTCCGGCACCCCTTCCGCC	1698
QY	241	ProGlnGlyAspArgGlyGlyGlyGlyPProGlnGlyGlyMetArgGlyGlyGlyGlyLeu	260
Db	1699	CCGGGTGGTGTCTGGCAGAGGTGGCCCTGTGGCATGCGGGGAGGAGAGGTGGCTC	1758
QY	261	MetAspArgGlyGlyGlyPProGlnGlyGlyMetPheArgGlyGlyGlyArgGlyGlyGly	280
Db	1759	ATGATAGTGTGTGTCTCCGGTGGAAATGTTCAGAGTGTCCGTGTGTGAAGACAGAGTGC	1818
QY	281	PheArgGlyGlyGlyArgGlyMetAspArgGlyGlyPheGlnGlyGlyGlyArgArgGlyPro	300
Db	1819	TTCCTGTGTGGCCGGGCGCATGAGCCGAGGTGGCTTGTGTGGAGGAACAGAGTGGCCT	1878
QY	301	GlyGlyProProGlyPProLeuMetGlnGlyMetGlyGlyGlyArgArgGlyGlyGlyGly	320
Db	1879	GGGGGGCCCCCTGGACCTTTGATGGAACAATGGAGAAAGAGAGGACGTTGGAGGA	1938
QY	321	ProGlnGlyMetAspArgGlyGlyGlnGlyArgGlnGlyArgArgArgArgArgArgArg	338
Db	1939	CCTGGAAAAATGATTAAGCGAGACCCGTACAGAGCGCAGAGATCGGCCCTAC	1992
RESULT 6			
LOCUS	AR080100	AR080100	2371 bp
DEFINITION	Sequence 1 from patent US 5968734.	LINEAR	PAT 31-AUG-2000
ACCESSION	AR080100		
VERSION	AR080100.1	GI:10006835	
KEYWORDS			
SOURCE	Unknown.		
ORGANISM	Unknown.		
REFERENCE	1 (bases 1 to 2371)		
AUTHORS	Aurias,A., Delacarte,O., Desmaze,C., Melot,T., Peter,M., Ploucasel,B., Thomas,G. and Zucman,J.		
TITLE	Nucleic acid corresponding to a gene of chromosome 22 involved in recurrent chromosomal translocations associated with the development of cancerous tumors, and nucleic acids of fusion resulting from said translocations		
JOURNAL	Patent: US 5968734-A 1 19-OCT-1999;		
FEATURES	Location/Qualifiers		
SOURCE	1..2371		
	/organism="Unknown"		
	/mol_type="unassigned DNA"		
ORIGIN			
Alignment Scores:			
Pred. No.:	3 56-47	Length:	2371
Score:	1922.00	Matches:	338
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	6	Gaps:	0
US-10-791-017A-2_COPY_319_656 (1-338) x AR080100 (1-2371)			
QY	1	GlyGlyArgGlyGlyMetGlySerAlaGlyGlyGlyGlyPheAsnGlyPProGlyGly	20
Db	979	GGAAGACCGCGGTGAATGCGCAGCGCTGGAGAGCGAGAGTGGCTTCATTAAGCCGTGGGA	1038
QY	21	PrometAspGlnGlyPProAspLeuAspLeuGlyPProProValAspPProAspGlnAspSer	40
Db	1039	CCCATGTAGTGAAGACCAAGATCTTGATCTTGGGCCCTCTGTAGATCCAGATGAAGACTCT	1098
QY	41	AspAsnSerAlaIleTyrValGlnGlyLeuAsnAspSerValThrLeuAspLeuAla	60
Db	1099	GACCAACGTGTGAATTTATGTACAAGAGATTAANTACAGTGTGACTCTTAGAGATCTGGCA	1158

QY	201	AlaGlyAspTrpGlnCysProAsnProGlyCysGlyAsnGlnAsnPheAlaTrpArgThr	220
Db	1598	GCTGAGACTGCGACAGTGTCCCAATCCGGATTGTGGAAACCAAGACTTCGTGCTGAGAAACA	1657
QY	221	GlucYsaAsnGlnCysIlyBaIaProIlyAsProGlnGlyPheIeuProProPhePro	240
Db	1658	GAGTGCACACAGTGTAAAGGCCCAAGCCTGAAAGGCTTCCTCCGCGCACCTTTCCGCGCC	1717
QY	241	ProGlyIlyIlyAspArgGlyIlyArgGlyIlyProGlyIlyMetArgGlyIlyIlyArgGlyIlyIleu	260
Db	1718	CCGGGTGTGTATCGTGGCAGAGGTGGCCCTGTGTGTGCATGCGGAGAGAAAGGTGGCCTC	1777
QY	261	MetAspArgGlyIlyIlyProGlyIlyIlyMetPheArgGlyIlyIlyArgGlyIlyIlyAspArgGlyIlyIly	280
Db	1778	ATGATCTGTGTGTGTCCTCCGTGTGAATGTCTTCAAGAGTGGCCCTGTGTGAAGACAGAGCTGGC	1837
QY	281	PheArgIlyGlyIlyArgGlyIlyMetAspArgGlyIlyIlyPheGlyIlyIlyIlyArgArgGlyIlyPro	300
Db	1838	TTCCGTGTGGCCGGGCGATGGAACGAGTGGCTTTGTGTGAAGAAAGCAGGTGGCCCT	1897
QY	301	GlyIlyIlyProProGlyIlyProIleuMetGlnIleuMetGlyIlyIlyArgArgGlyIlyIlyArgGlyIly	320
Db	1898	GCGGGGCCCTCCGGACTTTGATGTGAACAGTGGAGGAAAGAGGAGGACGTGTGAGGA	1957
QY	321	ProGlyIlyIlyMetAspIlyGlyIlyIlyIlyIlyArgGlnIlyIlyArgAspArgProIlyIly	338
Db	1958	CCTGGAATAATGAGATTAAGGCGAGACCCGTCTCAGAGCCGACAGATCGGCCCTTAC	2011

RESULT 8			
AX411125	AX411125	2390 bp	DNA
LOCUS			linear
DEFINITION	Sequence 3772 from Patent WO0229103.		PAT 14-JUN-2002

SOURCE	Homo sapiens (human)
...	...

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

REFERENCE
1
Alvares, C., Horne, D., Peres-da-Silva, S. and Vockley, J.G.
AUTHORS
TITLE
Gene expression profiles in liver cancer
JOURNAL
Patent: WO 0229103-A 3772 11-APR-2002;

FEATURES	location/Qualifiers
source	1. 2390
	/organism="Homo sapiens"
	/mol_type="unassigned DNA"
	/db_xref="taxon:9606"
	/note="EMBL/GenBank Accession No. X66899"

ORIGIN

Alignment Scores:	
Pred. No.:	3,52e-47
Score:	1972.00
Percent Similarity:	100.00%
Best Local Similarity:	100.00%
Query Match:	100.00%
DB:	6
Length:	2390
Matches:	338
Conservative:	0
Mismatches:	0
Indels:	0
Gaps:	0

US-10-791-017A-2_COPY_319_656 (1-338) x AX411125 (1-2390)

Qy 1 G1G1yArG1yG1yMeG1ySeA1aG1y1uArG1yG1yPheAsnLybProG1y 200
Db 998 GGAAGACCCGGTGAATGGCAGCCCTGGAGAGCGAGGTGGCTTCATTAAGCCTGGTGGGA 1000

QY 21 prometasp[glu]yproasp[leu]asp[leu]g[ly]proprovalasp[pro]asp[glu]asp[ser] 40

Db 1058 CCCATGGATGAAGGACCAGATCTTGATCTAGGCCCTCCTGTAGATCCAGATGAAGACTCT 11

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

Db 1118 GACACAGTGCATTATGTACAAGATTAAATGACAGTGTGACTCTAGATGATCTGGCA 1177

01 ABPBEFIEHLYBOLNYBOLYVALVADLYBNULCHNULYBAGYANOTYDANLZONCZACZC

81 HistIeTYrLeuAspIysGIuThrGIyIysProIysGIyAspAlaThrValSerTYrGIu 100

Db 1238 CACATCTACCTGACCAAGGAACAGGAAGCCAAAGGCGATGCCACAGTGTCTATGAA 1297

[illegible]

121 Ival,euI,vsVa] SerI,euA] aArcl,vsI,vsProP,romet,AsnSerI,etA,rgI,vgI,yl,eu 140

Db 1358 AACCTTAACTCTCCCTTCTCGGAGAAGCCTCCATGACACAGTATCGGGGTGTCG 1417

QY	141	ProProArgGluGluArgGlyMetProProProLeuArgGlyProGlyGlyProGly	160

[illegible]

Db 1478 GGTCCTGGGGGACCCATGGGTGCATGGAGCCGTGAGGAGATAGAGAGGCTTCCCT 1537

181 ProArgGlyProArgGlySerArgGlyAsnProSerGlyGlyGlyAsnValGlnHisArg 200

[illegible]

Db 1598 GCTGAGACTGGCAGTGTCCCAATCCGGGTTGTGGAAACGAACTTGGCTGGAGACA 1657

221 GluCysAsnGlnCysLysAlaProLysProGluGlyPheLeuProProProPheProPro 240

[illegible]

Db 1718 CCGGCTGGTGAATCGTCGACAGGTGGCCCTGTCGATCCGGGAGGAGAGGTGGCCTC 1777

261 MetAspArgGlyGlyProGlyMetPheArgGlyGlyArgGlyGlyAspArgGlyGly 280

Db
1/18 ATGGATCGTGGTGGTCCCGGTGGAAAGTTCAGTGTGGCCGATGGAGTCAAGGATGGC 203

1838 TTCCGTGCTGCGGGCATGACCGAGGTGCTTTGCTGAGGACGACGAGCTGGCCCT 189

301 GlyGlyProProGlyProLeuMetGluGlnMetGlyArgArgGlyArgGlyGly 320

Db 1898 GGGGGCCCCCTGGACCTTGTATGGAACAGATGGGAGAGAAGAGAGAGCTGGAGGA 195

1068 CCTCCAAATATCCGTAAGCCGACCCGTACAGAGCGAGAGATCGACCTAC 2011

RESULT 9

LOCUS	HSEWS	2390 bp	mRNA	linear	PR1 28-JUN-1995
DEFINITION	H. sapiens EMS mRNA.				

ACCESSION	A00000
VERSION	X66899.1
	GI:547565

SOURCE	
ORGANISM	Homo sapiens (human)
	Homo sapiens

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

AUTHORS
Bergallie, O., Zucman, O., Frougaber, B., Desmazière, C., Hecquet, J.,
Peter, M., Kovar, H., Joubert, I., de Jong, P., Rouleau, G., Aurias, A.

AUTHORS Isogai, T., Yamamoto, J., Nishikawa, T., Isono, Y., Sugiyama, T.,

TITLE Otsuki, T., Makamatsu, A., Ishii, S., Nagai, K. and Irie, R.

JOURNAL Full-length human cDNA

Patent: EP 1447413-A 952 18-AUG-2004;

Research Association for Biotechnology (JP)

FEATURES
source 1. 2026
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"

ORIGIN

Alignment Scores:

Pred. No.: 7 4e-47 Length: 2026
Score: 1909.00 Matches: 337
Percent Similarity: 99.70% Conservative: 0
Best Local Similarity: 99.70% Mismatches: 1
Query Match: 99.32% Indels: 0
Gaps: 0

US-10-791-017a-2_COPY_319_656 (1-338) x CQ850483 (1-2026)

QY 1 G1YGLYARGLYGLYMETGLYSEALAG1YGLUARG1YGLYPHEANLYSProG1Y 20
DB 836 GGAGGACGGGTGGATGGGCAAGGCTGGAGAGAGAGTGGCTTCAATTAAGCCTGGTGA 895
QY 21 ProMetAspGluG1YProAspLeuAspLeuG1YProProValAspProAspGluAspSer 40
DB 896 CCAATGATGATGAGACCAATCTTGATCTAGCCCACTTGAATCCAGATGAGAGACTCT 955
QY 41 AspaAspSerAlaIleTYrValG1NG1YLeuAspAspSerValThrLeuAspAspLeuAla 60
DB 956 GACACAGCTGCAATTTATGTCACAGATTTAAATACAGTGTGACTTCAATGATGATGCA 1015
QY 61 AspPhePheLeuG1YGLYValValValMetAsnLYsArgThrG1YGLNProMetIle 80
DB 1016 GACTTCTTAAAGCAGTGTGGGTGTGTTAAGATGAAACAAGAGAACTGGGCAACCATGATC 1075
QY 81 HisIleTYrLeuAspLYsG1YThrG1YLYsProLYsG1YAspAlaThrValSerTYrGlu 100
DB 1076 CACATCTAAGCTGACCAAGAAACGAAAGCCCAAGGCAATGCAAGTGTCTTAAGAA 1135
QY 101 AspProProThralValAlaValAlaValG1UTrPheAspG1YLYsAspPheG1YLYsSer 120
DB 1136 GACCCACCCACCTGCCAAGGCTGCGGTGATGTTTATGGAAGATTTTCAAGGAGAC 1195
QY 121 LysLeuLYsValSerLeuAlaArgLYsAspProProMetAspSerMetArgG1YLYsLeu 140
DB 1196 AAACCTTAAAGTCTCCCTTGTCTCGGAAAGAGCTTCAATGAACAGTATGCGGGTGTCTG 1255
QY 141 ProProArgG1YGLYARG1YMETProProProLeuArgG1YGLYProG1YGLYProG1Y 160
DB 1256 CCACCCCGTGAAGGCAAGAGATGCCACACCACTCCGTGAGAGTCCAGAGAGGCCCGCA 1315
QY 161 G1YProG1YGLYProMetG1YARGMetG1YGLYARG1YGLYAspArgG1YGLYPhePro 180
DB 1316 GGTCTCGGGGACCCCAAGGCTGCGATGAGAGGCCGTGAGAGATTAAGAGGCTTCCCT 1375
QY 181 ProArgG1YProArgG1YSerArgG1YAsnProSerG1YGLYAsnValG1NH1Arg 200
DB 1376 CCAAGAGAGACCCCGGGGTTCGCGAGGAGAACCCCTCTGAGAGAGAAAGTCCAGACCGA 1435
QY 201 AlaG1YAspTrpG1YArgProAsnProG1YCYsG1YAsnG1NAsnPheAlaTrpArgThr 220
DB 1436 GCTGGAGAACCTGCAAGTGTCCCAATCCGGGTGTGAAACCAAGAACTTCCCTGAGAGAA 1495
QY 221 GluCYsAsnG1NAsnValAlaProLYsProG1YGLYPheLeuProProProPheProPro 240
DB 1496 GAGTGCACACAGTAAAGGCCCAAGAGCTTAAGAGCTTCTCCGCAACCTTTCCGCCCC 1555
QY 241 ProG1YGLYAspArgG1YARG1YGLYProG1YGLYMetArgG1YGLYARG1YGLYLeu 260
DB 1556 CCGGTGTGTATCTGGCAAGAGTGTGCTGTGTGATGCGGGAGAGAAAGGTGGCTC 1615

QY 261 MetAspArgG1YGLYProG1YGLYMetPheArgG1YGLYARG1YGLYAspArgG1YGLY 280

DB 1616 ATGATGCTGTGTGTGTCCCGATGGAATTTCAAGAGTGGCCGTGTGAGACAGAGGTGGC 1675

QY 281 PheArgG1YGLYARG1YMETAspArgG1YGLYPheG1YGLYGLYARG1YGLYPro 300

DB 1676 TTCGGTGTGTGCGGGGATGAGACCGAGGTGCTTGTGTGAGAGAAAGACAGAGTGGCCCT 1735

QY 301 G1YGLYProProG1YProLeuMetG1YGLNMetG1YGLYARGArgG1YGLYARG1YGLY 320

DB 1736 GGGGGGCCCCCTGAGACTTGTATGAAACAGATGGAGAGAAAGAGAGACACCTGAGGA 1795

QY 321 ProG1YLYsMetAspLYsG1YGLNHisArgG1NG1YARGArgAspArgProTYr 338

DB 1796 CTTGAAAAATGATTAAGCCAGCACCTCAGAGACCGAGAGATCGGCCCTAC 1849

RESULT 11

AK127624

LOCUS

DEFINITION

AK127624

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

1

Oshima, A., Takahashi, Fujii, A., Tanase, T., Inoue, N., Takeuchi, K.,

Arita, M., Muraashiro, K., Yuuki, H., Hara, H., Sugiyama, T., Irie, R.,

Otsuki, T., Sato, H., Makamatsu, A., Ishii, S., Yamamoto, J., Isono, Y.,

Kawai, H., O., Saito, K., Nishikawa, T., Kimura, K., Yamashita, H.,

Matsuo, K., Nakamura, Y., Sekine, M., Kikuchi, H., Kanda, K.,

Wagatsuna, M., Muraikawa, K., Kanehori, K., Sugiyama, A., Kawakami, B.,

Suzuki, Y., Sugano, S., Nagahara, K., Masuho, Y., Nagai, K. and

Isogai, T.

NEBO human cDNA sequencing project

Unpublished

2 (bases 1 to 2026)

Isogai, T. and Yamamoto, J.

Direct Submission

Submitted (15-JUL-2003) Takeo Isogai, FUJ Project (HRI Team); 2-6-7

Kazusa-Kamatari, Kisarazu, Chiba 292-0818, Japan

(E-mail: genomics@hri.co.jp, Tel: 81-438-52-3975, Fax: 81-438-52-3986)

NEBO human cDNA sequencing project supported by Ministry of

Economy, Trade and Industry of Japan; cDNA full insert sequencing:

Research Association for Biotechnology (RAB); cDNA library

construction: Helix Research Institute (HRI) (supported by Japan

Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB,

HRI, and Biotechnology Center, National Institute of Technology and

Evaluation; clone selection for full insert sequencing: HRI and

RAB; annotation: Reverse Proteomics Research Institute, HRI and

RAB.

Location/Qualifiers

1. 2026

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="HCASM2008154"

/cell_type="coronary artery smooth muscle cells (HCASMC)"

/note="cloning vector: pME185FL3-primary culture, coronary

artery smooth muscle cells"

ORIGIN

Alignment Scores:

Pred. No.: 7 4e-47 Length: 2026

Score: 1909.00 Matches: 337

Percent Similarity: 99.70% Conservative: 0

Best Local Similarity: 99.70% Mismatches: 1

Query Match: 99.32% Indels: 0

DB: 9 Gaps: 0

US-10-791-017a-2_COPY_319_656 (1-338) x AK127624 (1-2026)

QY 1 G1YGLYARGLYGLYMETGLYSEALAGLYGLYARGLYGLYPhEaenLYsProGLYGLY 20

DB 836 GAGAGACGGCGTGAATGGGCAAGCGCTGGAGAGAGGTGCTTCAATAGCCTGGTGA 895

QY 21 ProMetAspGLUGLYProAbpLeuApLeuGLYProProValAspProApGLUAspSer 40

DB 896 CCATGATGATGAAGACCAAGATCTTGATCTAGCGCCACCTGTAGATCCAGATGAAGACTCT 955

QY 41 AspAsnSerAlaIleTyrValGlnGlyLeuAsnAspSerValThreLeuAspAlaLeuAla 60

DB 956 GACAAAGATGCAATTTATGTAACAAGATTAAGACATGATGACTGATGATGATGATGATGATG 1015

QY 61 AspPhePheLYsGLYValValLYsMetAsnLYsAspThrGLYGLYProMetIle 80

DB 1016 GACTTCTTAAAGCAGTGGGTTGTTAAGATGAACAAGAACTGGGCAACCATGATC 1075

QY 81 HisIleTyrLeuAspLYsGLYValValLYsProLYsAspAlaThrValSerTyrGlu 100

DB 1076 CACATCTACCTGACCAAGAAACAGAAAGCCCAAGGCGATGCCACAGTGTCTATGAA 1135

QY 101 AspProThrAlaValAlaValGluTyrPheAspGLYLYsAspPheGLYGLYSer 120

DB 1136 GACCCACCACTGCACAGGCTGCCGTGAAATGTTGATGGAAGATTTTCAAGGAGC 1195

QY 121 LYsLeuLYsValSerLeuAlaTyrLYsProPheAsnSerMetAspGLYGLYLeu 140

DB 1196 AAACCTTAAGTCTCCCTGCTCCGAAGAACCTTCAATGAACATGATGCGGGTGGCTG 1255

QY 141 ProProArgGLUGLYARGLYMETProProProLeuArgLYGLYProGLYGLYProGLY 160

DB 1256 CCAACCCCGTGAAGGAGAGCATGCCACCACTCGTGAAGTCCAGAGGCCCAAGCA 1315

QY 161 GLYProGLYGLYProMetGLYARGLYGLYARGLYGLYAspArgGLYGLYPhePro 180

DB 1316 GGTCTGGGGGACCCATGGTCCCATGGAGGCGGTGGAAGATGAAGAGGCTTCCCT 1375

QY 181 ProArgGLYProArgLYsSerArgLYsAspProSerGLYGLYLYsAsnValGlnHisArg 200

DB 1376 CCAAGAGAACCCCGGGGTTCCCAAGGAAACCTCTGAGAGAGAAACGTCAGACCA 1435

QY 201 AlaGLYAspTyrGlnCysProAsnProGLYLYsGLYAsnGlnAsnPheAlaTyrPArgTyr 220

DB 1436 GCTGGAGACCTGCAAGTCCCATCCGGGTGTGGAACCAAGAACTTCCGCTGAGAAC 1495

QY 221 GluCysAsnGlnCysLeuValAlaProLYsProGLYGLYPheLeuProProPheProPro 240

DB 1496 GAGTGCACCAAGTGAAGGCCCAAGGCTTGAAGGCTTCCCGCACCTTTCGCCCC 1555

QY 241 ProGLYGLYAspArgLYsArgLYsGLYProGLYGLYMETArgLYsGLYArgGLYGLYLeu 260

DB 1556 CCGGATGGTGAATGTCGAGAGGTCCTGATGCAATGGGGGAGAGAGAGTGGCTC 1615

QY 261 MetAspArgGLYGLYProGLYGLYMETPheArgGLYGLYARGLYGLYAspArgGLYGLY 280

DB 1616 ATGATGATGATGATGATCCCGGTGAATGTTCAAGAGTGGCGGTGTGAGACAGAGGTGG 1675

QY 281 PheArgGLYGLYARGLYMETAspArgGLYGLYPheGLYGLYLYsArgArgGLYGLYPro 300

DB 1676 TTCCTGGTGGCGGGGCAATGACCAAGGTCCTTGGTGAAGAAACAGAGTGGCCCT 1735

QY 301 GLYGLYProProGLYProLeuMetGLYGLYMETGLYGLYARGLYGLYArgGLYGLYArgGLY 320

DB 1736 GGGGGGCCCCCTGACCTTGAATGACAGATGGAGAGAAAGAGAGAGACCTGAGAGA 1795

QY 321 ProGLYLYsMetAspLYsGLYGLYHisArgGLYGLYArgArgArgProTyr 338

DB 1796 CCTGGAATAATGATTAAGCGAGCACCGTCAAGAGCGCAGAGATCGGCCCTTAC 1849

RESULT 12

CQ730352

LOCUS CQ730352 2326 bp DNA linear PAT 03-FEB-2004

DEFINITION Sequence 16286 from Patent WO02068579.

ACCESSION CQ730352

VERSION CQ730352.1 GI:42303963

KEYWORDS

SOURCE

ORGANISM

Homo sapiens (human)

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE

1

VENTER, C.J., ADAMS, M.C., LI, P.W. and MYERS, E.W.

1998

Genome-wide analysis of human gene expression using complementary DNA arrays

Genome Res. 8: 116-124

JOURNAL

Patent: WO 02068579-A 16286 06-SEP-2002;

PE Corporation

FEATURES

source

1..2326

location/Qualifiers

organism="Homo sapiens"

molecule="unassigned DNA"

db_xref="taxon:9606"

ORIGIN

Alignment Scores:

Pred. No.: 8.11e-47 Length: 2326

Score: 1909.00 Matches: 338

Percent Similarity: 99.71% Conservative: 0

Best Local Similarity: 99.71% Mismatches: 0

Query Match: 99.32% Indels: 1

DB: 6 Gaps: 0

US-10-791-017a-2_COPY_319_656 (1-338) x CQ730352 (1-2326)

QY 1 G1YGLYARGLYGLYMETGLYSEALAGLYGLYARGLYGLYPhEaenLYsProGLYGLY 20

DB 950 GAGAGACGGCGTGAATGGGCAAGCGCTGGAGAGAGGTGCTTCAATAGCCTGGTGA 1009

QY 21 ProMetAspGLUGLYProAbpLeuApLeuGLYProProValAspProApGLUAspSer 40

DB 1010 CCATGATGATGAAGACCAAGATCTTGATCTAGGCCCACTGATATCAAGTGAAGACTCT 1069

QY 41 AspAsnSerAlaIleTyrValGlnGlyLeuAsnAspSerValThreLeuAspAlaLeuAla 60

DB 1070 GACAAAGATGCAATTTATGTAACAAGATTAAGACATGATGATGATGATGATGATGATG 1129

QY 61 AspPhePheLYsGLYValValLYsMetAsnLYsAspThrGLYGLYProMetIle 80

DB 1130 GACTTCTTAAAGCAGTGGGTTGTTAAGATGAACAAGAACTGGGCAACCATGATC 1189

QY 81 HisIleTyrLeuAspLYsGLYValValLYsProLYsAspAlaThrValSerTyrGlu 100

DB 1190 CACATCTACCTGACCAAGAAACAGAAAGCCCAAGGCGATGCCACAGTGTCTATGAA 1249

QY 101 AspProThrAlaValAlaValGluTyrPheAspGLYLYsAspPheGLYGLYSer 120

DB 1250 GACCCACCACTGCACAGGCTGCCGTGAAATGTTGAGGAGAAATTTTCAAGGAGC 1309

QY 121 LYsLeuLYsValSerLeuAlaTyrLYsProPheAsnSerMetAspGLYGLYLeu 140

DB 1310 AAACCTTAAGTCTCCCTGCTCCGAAGAACCTTCAATGAACATGATGCGGGTGGCTG 1369

QY 141 ProProArgGLUGLYARGLYMETProProProLeuArgLYGLYProGLYGLYProGLY 160

DB 1370 CCAACCCCGTGAAGGAGAGCATGCCACCACTCGTGAAGTCCAGAGGCCCAAGCA 1429

QY 161 GLYProGLYGLYProMetGLYARGLYGLYARGLYGLYAspArgGLYGLYPhePro 180

DB 1430 GGTCTGGGGGACCCATGGTCCCATGGAGGCGGTGGAAGATGAAGAGGCTTCCCT 1489

QY 181 ProArgGLYProArgLYsSerArgLYsAspProSerGLYGLYLYsAsnValGlnHisArg 200

DB 1490 CCAAGAGAACCCCGGGTTCGAGAGAACCTCTTGAAGAGAGAAAGTCCAGACCA 1549

QY 21 ProMetAapgluclgylProaapleuapleuclgylProProVa1aapProapgluapSer 40
 Db 1018 CCATGATGAAGAGACCAAGCTCTTGAATCTAGGCCCACTTGAATCCAGATGAAGACTCT 1077
 QY 41 AapAasSerAla1eTyrVal1gnglyLeuAasAasSerVal1ThrLeuAapAapleuAla 60
 Db 1078 GACAAACAGTCAATTTATGTAACAGATTTAAATGACAGTGTGACTCTTGAATGATCTGGCA 1137
 QY 61 AapPhePheylsglncysglVal1vallyMetAasNylsargThrglyglnPrometIle 80
 Db 1138 GACTTCTTTAAGCAGGTGGGGTGTGTTAATGATGAACAAGAGAACTGGGCAACCATGATC 1197
 QY 81 Hle1leTyrLeuAaplysgluthrglyLysProlysglyAapAlaThrVal1SerTyrGlu 100
 Db 1198 CACATCTAAGCTGACAGAAACAGAAACAGAAACAGAAACAGAAACAGAAACAGAAACAGAA 1257
 QY 101 AapProProthral1alysAla1aValGluTyrPheAapgl1LysAapPheglnlySer 120
 Db 1258 GACCCACACCTGCAAGGCTGCGGTGGAAATGGTTTATGGAAAGATTTTCAAGGAGC 1317
 QY 121 LysLeuLysVal1SerLeuAlaarglyLysProProMetAasSerMetAarglyLysLeu 140
 Db 1318 AAACCTTAAAGTCTCCCTGCTCGGAAGAAGCCTTCAATGAACAGTATGCGGGGTGCTG 1377
 QY 141 ProProAarglyglu1yarglyMetProProProleuAarglygl1yProgl1yLysProgl 160
 Db 1378 CCAACCCCGTGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1437
 QY 161 Gl1yProgl1ygl1yProMetgl1yargMetgl1ygl1yAarglygl1yAapAarglygl1yPhePro 180
 Db 1438 GGTCTCTGGGGAGCCCATGAGTGGATGGAGGCGGTGGAGAGATGAAGAGAGGCTTCCCT 1497
 QY 181 ProAarglyProAarglyLysAarglyLysAasProSerSer1ygl1yAanVal1GlnH1sarg 200
 Db 1498 CCAAGAGGAGCCCGGGGTTCCCGAGGAGAACCCCTCTGAGAGAGAGAAACGTCAGACCGA 1557
 QY 201 AlAGlyAapTyrPglncysPProAanProgl1yCysgl1yAanGlnAanPheAlaTTPAargThr 220
 Db 1558 GCTGAGAGACTGCGAGTGTCCCAATCCGGGTGTGTAACCAAGAACTTCGCTCGAGAGCA 1617
 QY 221 GluCyAanGlnCysLysAla1aProlyPProgl1yPheLeuProProPheProPro 240
 Db 1618 GAATGCAACCAAGTGAAGGCCCAAGCCGAAAGGCTCTCCCGCAACCTTTCGCGCC 1677
 QY 241 Progl1yLysAapAarglyLysAarglyLysProgl1yLysAarglyLysAarglyLysLeu 260
 Db 1678 CCGGCTGTGATCTGTGCAAGAGTGGCCCTGTGTGCAATGCGGGAGAGAAAGGTGGCTTC 1737
 QY 261 MetAapAarglygl1yProgl1yMetPheAarglygl1yAarglygl1yAapAarglygl1y 280
 Db 1738 ATGATGATGTGTGTGCTCCGCTGGAATTTTCAAGAGTGGCGGTGTGAGACACAGAGGTGC 1797
 QY 281 PheAarglygl1yAarglyMetAapAarglygl1yPhegl1ygl1yAargAarglygl1yPro 300
 Db 1798 TTCGGTGTGTGCGGGGCAATGACCGAGGTGGCTTGTGTGTGTGTGTGTGTGTGTGTGTGT 1857
 QY 301 Gl1yLysProProgl1yProleuMetgl1yLysMetgl1yLysAarglygl1yAarglygl1y 320
 Db 1858 GGGGGGGCCCTGAGACTTGTGATGAGAACATGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1917
 QY 321 Progl1yLysMetAaplysglygl1yLysAarglygl1yAarglygl1yAapAargProTyr 338
 Db 1918 CCTGGAATAATGATTAAGCGAGCACTGTCAAGAGCGAGAGATCGGCGCTTAC 1971

RESULT 14

AK056309 2189 bp mRNA linear PRI 30-JAN-2004
 LOCUS Homo sapiens cDNA FLJ31747 f1s, clone NT2R12007377, highly similar
 DEFINITION to RNA-BINDING PROTEIN EMS.
 ACCESSION AK056309
 VERSION AK056309.1 GI:16551673
 KEYWORDS oligo capping; f1s (full insert sequence).

SOURCE

ORGANISM

REFERENCE

AUTHORS

Homo sapiens (human)
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

TITLE

JOURNAL

PUBMED

REFERENCE

1702039
 Nat. Genet. 36 (1), 40-45 (2004)

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

FEATURES

SOURCE

1. 2189
 Location/Qualifiers
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="NT2R12007377"
 /cell_line="NT2"
 /cell_type="teratocarcinoma"
 /clone_lib="NT2R12"

PGPMBEGPDLIDGPVDPDSDNSAIYVQGLNDSVTLIDLADFPKCGVVMNKR
GOMTHIYLDKTKGPKKGDATVSEYEDPTAKAIVEMWDGDFGSKLKVSLAKKPPM
NSMRGGLPRREGKMPPLRGPGGPGGPGGPGGPGGPGGPGGPGGPGGPGGPGG
SGGAVQHRADQCCPNPGCQNOFAFRTSCNCKAPKPGFLPPPPPGGGRGNG
PGGMRGGRGLMDRGPGGMPFRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRG
MEQMGRRGRGGRGPGKMDKEHROERRDRY"

ORIGIN

Alignment Scores:

Pred. No.:	8.59e-47	Length:	2189
Score:	1907.50	Matches:	337
Percent Similarity:	99.70%	Conservative:	0
Best Local Similarity:	99.70%	Mismatches:	0
Query Match:	99.25%	Indels:	1
DB:	9	Gaps:	1

US-10-791-017a-2_COPY_319_656 (1-338) x CR456490 (1-2189)

QY 1 G1YGLYARGLYGLYMETGLYSEALAGLYGLUARGLYGLYLPHEANLYSPROGLYGLY 20
DB 988 GGAGGACCGCGTGGAATGGC--GCTGAGAGCGAGTG3CTTCAATTAAGCCTGGTGGA 1044
QY 21 PrometAmpGLUGLYProAmpLeuAmpLeuGLYProProValAmpProAmpGLUAmpSer 40
DB 1045 CCCATGGATGAAGGACCAATCTTGATCTAGGCCCACTGTAGATCCAGATGAAGACTCT 1104
QY 41 AspAmpSerAlaIleTYValGLINGLYLeuAmpAmpSerValThreAmpAmpLeuAla 60
DB 1105 GACAAACAGTCCAATTATGTACAAAGATTAAAGACAGTGTGACTAGATGATCTGGCA 1164
QY 61 AspPhePheLeuGLYINGLYValValLYMeAmpAmpLYArgThrgLYGLINPMeTlle 80
DB 1165 GACTCTTTAAGCAGTGGGGTGTGTAAATGAACAAAGAACTGGGCAACCATGATC 1224
QY 81 HsILeTYLeuAmpLYSGIuThrgLYLYPProLYGLYAspAlaThrValSerTYRGlu 100
DB 1225 CACATCTACCTGAGCAAGAAACAGAAACCCAAAGCGATGCCACAGTGTCTATGAA 1284
QY 101 AspProProThraIaLYValaIaValGLUTPheAmpGLYLYAspPheGLINGLYSer 120
DB 1285 GACCCACCCCACTCCAAAGGCTGCCGTGGAATGGTTGATGGAAAGATTTCAGAGGAGC 1344
QY 121 LYLeuLYValSerLeuAlaArgLYValLYPProPheAmpSerMetArgGLYGLYLeu 140
DB 1345 AAACCTTAAGTCTCCCTTGCTCGAAGAAAGCTTCAATGAACAGTATGCCGGGTGTCTG 1404
QY 141 ProProArgGLIUGLYARGLYMetProProProLeuAmpGLYGLYProGLYGLYProGLY 160
DB 1405 CCAACCCCGTGAAGGCGAGGCAATGCCACCACTCCGTGAGGTCCAGAGGCCCAAGGA 1464
QY 161 G1YProGLYGLYProMetGLYArgMetGLYLYArgGLYGLYAspArgGLYLYPhePro 180
DB 1465 GGTCTCGGGGGAACCATGGGTGCGCATGGGAGGCCGTGAGAGATAGAGGAGGCTTCCT 1524
QY 181 ProArgGLYProArgGLYSerArgGLYAspProSerGLYGLYLYAsnValGINHISArg 200
DB 1525 CCAAGAGAACCCCGGGGTCTCCGAGGGAAACCTCTGAGAGAGAAACGTCCAGACCGGA 1584
QY 201 AlaGLYAspTPGInCYsProAmpProGLYCYeGLYAsnGLInAmpPheAlaTPArgThr 220
DB 1585 GCTGGAGACTGGCAAGTCTCCCAATCCGGGTGTGGAACCAAGAACTTCGCTGAGAGACA 1644
QY 221 G1UCyAAsnGLINCyAlYValaProLYSPROGLIUGLYPheLeuProProPheProPro 240
DB 1645 GAGTGCAACCAAGTAAAGGCCCAAGGCTTCTCCGCCCAACCTTCCGCCCC 1704
QY 241 ProGLYGLYAspArgGLYArgGLYGLYProGLYGLYMetArgGLYGLYArgGLYGLYLeu 260
DB 1705 CCGGGTGGTGTGTGGCAGAGGTGGCCCTGTGGCATGCGGGAGAGAAAGGTGGCTTC 1764
QY 261 MetAmpArgGLYGLYProGLYGLYMetPheArgGLYGLYArgGLYGLYAspArgGLYGLY 280
DB 1765 ATGGAATCGTGTGTCTCCCGGTGAATTTCAAGAGTGGCCGTGTGAGAGACAGAGGTGGC 1824

QY 281 PheArgGLYGLYARGLYMetAmpArgGLYGLYPheGLYGLYGLYArgArgGLYGLYPro 300
DB 1825 TTCGTGTGTGGCCGGGCGATGAGCCAGAGGTGGCTTTGTGTGAGAGAAAGAGAGGAGGCTTC 1884
QY 301 G1YGLYProProGLYProLeuMetGLIUGLYMetGLYGLYArgArgGLYGLYArgGLYGLY 320
DB 1885 GGAGGAGCCCTGACCTTGTATGAAACAGATGGAGAGAAAGAGAGAGAGAGAGAGAGAGAG 1944
QY 321 ProGLYLYMetAmpLYSGIUGLYLISArgGLINGLYArgArgAmpArgProTYR 338
DB 1945 CTTGAAAAATGATTAAGCGAGCAACCTTCAGAGAGCGCAGAGATGGCCCTTAC 1998

Search completed: February 21, 2005, 02:07:41
Job time : 3639.31 secs

Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - nucleic search, using frame_plus.p2n model

Run on: February 20, 2005, 16:18:03, Search time 427.771 Seconds
(without alignments)
4677.443 Million cell updates/sec

Title: US-10-791-017a-2_COPY_319_656
Perfect score: 1922
Sequence: 1 GGRGKMGSGRGRGFRKPKPG.....GGRGKMDKGRHQRDRRY 338

Scoring table:
BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODE=frame+.p2n.model -DEV=xlp
-Q/cg2.1/USPTO.spool.p/US10791017/runat.17022005.125806.22064/app.query.fasta_1.1358
-DB=N Geneseg.16dec04 -QFMT=fastap -SUFFIX=p2n.rng -MINMATCH=0.1 -LOOPEXT=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=humand0.cdt
-LIST=45 -DOCALLIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pct -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US10791017_@CGN.1.1.886_@runat.17022005.125806.22064 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -MARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N_Geneseg.16dec04:*
1: geneseg19808:*
2: geneseg19908:*
3: geneseg20008:*
4: geneseg20018:*
5: geneseg20018:*
6: geneseg20028:*
7: geneseg20028:*
8: geneseg20038:*
9: geneseg20038:*
10: geneseg20038:*
11: geneseg20038:*
12: geneseg20048:*
13: geneseg20048:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1922	100.0	1988	10	ADA53506
2	1922	100.0	2176	6	AA562262
3	1922	100.0	2371	2	AA050643
4	1922	100.0	2372	13	AD086032
5	1922	100.0	2388	5	AA570647

6	1922	100.0	2390	6	ABK64628	Abk64628 Human CDN
7	1922	100.0	2390	6	ABK64822	Abk64822 Human ben
8	1922	100.0	2390	6	ABN97274	Abn97274 Gene #377
9	1922	100.0	2390	13	AD516277	Ad516277 Human CDN
10	1922	100.0	2390	13	ACN40903	Acn40903 Tumour-as
11	1909	99.3	2026	13	ADR07446	Adr07446 Full leng
12	1888.5	98.3	2188	6	AB199383	Ab199383 Mouse isc
13	1810	94.2	1807	13	ADP56333	Adp56333 Human PRO
14	1777	92.5	2273	6	AA562623	Aa562623 cDNA sequ
15	1742	90.6	3309	2	AA050646	Aa050646 Human Ews
16	1619	84.2	2177	13	ACN37537	Acn37537 Tumour-as
17	1385	72.1	1462	4	AA126725	Aa126725 Human bre
18	1181	61.4	2299	11	ACN86697	Acn86697 Breast ca
19	961	50.0	350	9	ACH40478	Ach40478 Human foe
20	878	45.7	1822	6	AB573275	Ab573275 DNA encod
21	878	45.7	1822	13	ADR14648	Adr14648 Human NF-
22	878	45.7	1824	6	AB573272	Ab573272 DNA encod
23	878	45.7	1824	6	ABK64149	Abk64149 Human CDN
24	878	45.7	1824	12	AD126116	Ad126116 Human CDN
25	878	45.7	1824	12	AD126116	Ad126116 Human CDN
26	878	45.7	1824	13	ACN39977	Acn39977 Tumour-as
27	878	45.7	1939	12	ADR77157	Adr77157 Human CDN
28	878	45.7	2179	4	AAK51488	Aak51488 Human pol
29	878	45.7	2209	4	AAK51489	Aak51489 Human pol
30	878	45.7	2213	4	AAK52472	Aak52472 Human pol
31	878	45.7	2213	4	AAK52473	Aak52473 Human pol
32	873.5	45.4	1832	12	AD126112	Ad126112 Human CDN
33	809.5	42.1	568	9	ACH22345	Ach22345 Human adu
34	750	39.0	2144	12	AD126206	Ad126206 Human CDN
35	750	39.0	2153	6	ABK64369	Abk64369 Human CDN
36	750	39.0	2153	12	AD126208	Ad126208 Human CDN
37	749	39.0	2126	3	AA64661	Aa64661 cDNA enco
38	733	38.1	471	2	AAV86051	Aav86051 EST clone
39	718.5	37.6	545	3	AACT8275	Aac78275 Human can
40	703	36.6	1610	11	AD12631	Ad12631 Human dis
41	622	32.4	665	13	AD054544	Ad054544 Novel can
42	616.5	32.1	1619	4	AB106387	Ab106387 Drosophila
43	593	30.9	361	8	ABX38686	Abx38686 Bovine ES
44	572.5	29.8	550	6	AB199315	Ab199315 Mouse isc
45	530.5	27.6	6664	4	AB106386	Ab106386 Drosophila

ALIGNMENTS

RESULT 1	ADA53506	standard; cDNA; 1988 BP.
ID	ADA53506	
AC	ADA53506	
XX		
XX		
DT	20-NOV-2003	(first entry)
XX		
DE	Human coding sequence, SEQ ID 1074.	
XX		
XX		
OS	Homo sapiens.	
XX		
PN	EP1293569-A2.	
XX		
PD	19-MAR-2003.	
XX		
XX		
PF	21-MAR-2002; 2002EP-00006586.	
XX		
PR	14-SEP-2001; 2001JP-00328381.	
XX		
PR	24-JUN-2002; 2002US-0350435P.	
XX		
PA	(HELI-) HELIX RES INST.	
XX		
XX	(REAS-) RES ASSOC BIOTECHNOLOGY.	
PI	Isogai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S,	

PI Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I;
PI Seki N, Yoshikawa T, Otsuka M, Nagahari K, Masuo Y;
XX MPI; 2003-395539/38.
DR P-PSDB; ADA55145.
XX
PT New polynucleotides encoding full-length polypeptides, e.g. secretory
PT and/or membrane proteins, useful for developing medicines for diseases in
PT which the gene is involved, or as target molecules for gene therapy.
XX
PS Claim 1; SEQ ID NO 1074; 205bp; English.
XX
CC The present invention relates to novel human secretory or membrane
CC proteins (ADA54072-ADA55710) and their coding sequences (ADA52433-
CC ADA54071). The coding sequences are useful in the gene therapy of
CC diseases caused by abnormalities of the proteins, e.g. cancer,
CC inflammatory diseases, osteoporosis or neurological disease.
XX
SQ Sequence 1988 BP; 519 A; 504 C; 584 G; 381 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 1,89e-60 Length: 1988
Score: 1922.00 Matches: 338
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: Gaps: 0

US-10-791-017a-2_copy_319_656 (1-338) x ADA53506 (1-1988)

QY 1 G1YGLYARGLYGLYMETGLYSERLAAG1YGLUARG1YGLYPHEANLYSPROGLYGLY 20
DB 798 GAGAGACGCGGTGATGGCAGCGCTGGAGAGCGAGGCTTCAATTAAGCTTGATGA 857
QY 21 PROMETASPLUG1YPROASPLEUASPLUG1YPROVALIAPROASPLUG1YUASPSER 40
DB 858 CCATGGATGAGAGACAGATCTTGAATCTAGGCCACCTGTAGTCCAGTGAAGACTCT 917
QY 41 ASPANSERALAIETRYVALINGLYLEUASNPASERVALTHIRLEUASPLUALLA 60
DB 918 GACAAACAGTGCAATTATGATGACAGATTAAATGACAGTGTGACTAGTGTGCGCA 977
QY 61 ASPPHEPHELYG1YVCSGLYVALIALLYMETASPLUARGTHRG1YGLINPROMETILE 80
DB 978 GACTTCTTTAAGCAGTGTGGGTTGTTAATGATGAACAGAACTGGCAACCAATGATC 1037
QY 81 H1S1ETRYLEUASPLUG1YURHGLYLYSPROLYSGLYASPLATHVALISERTYGLU 100
DB 1038 CACATCTACCTGGACAGAAACAGAAACCCAAAGCGATGCCACAGTGTCTTAGAA 1097
QY 101 ASPPROPTHRLALYVALAALAVAGLUTTPHEASPLIYLYASAPPHREG1YGLYSER 120
DB 1098 GACCCACCCACCTCCAGAGGTGCGGTGGAATGTTGATGGAAAGATTTTCAAGGAGC 1157
QY 121 LYELEUVALSERLEUALAARGLYLYSPROPROMETASNPASERMETARG1YGLYLEU 140
DB 1158 AAACCTTAAGTCTCCCTGCTCGAAGAACCTTCAATGAACGATATCGGGTGTCTG 1217
QY 141 PROPROARG1YGLYARG1YMETPROPROPROLEUARG1YGLYPROGLYGLY 160
DB 1218 CCAACCCGTAAGGAGAGAGCATGCGCACCACTCGTGTGAGAGGTCCAGAGGCCAGGA 1277
QY 161 GLYPROGLYGLYPROMETGLYARGMETGLYLYARG1YGLYASPLARG1YGLYPHEPRO 180
DB 1278 GGTCTCGGGGAGCCATGGGTGCGATGGAGGCGGTGAGAGATAGAGAGGCTTCCCT 1337
QY 181 PROARG1YPROARG1YSEARG1YASNPASERGLYGLYGLYASPLVALGINH1SARG 200
DB 1338 CCAAGAGACCCCGGGGTTCCCAAGGAAACCTCTTGAAGAGAAACGTCACACCGCA 1397
QY 201 ALAAG1YASPLTPGLYCYSPROASNPARG1YCSGLYASNG1YASNPHEALATPARGTHR 220
DB 1398 GCTGAGAGACTGGAGTGTCCCATCCGGGTGTGGAAACAGAACTTCGCTGAGAGACA 1457

QY 221 GLUCYBAENG1CYBULEYENLAEROLYSPROGLUG1YPHELEUPROPROPHOPROPRO 240
DB 1458 GAGTGCACCAAGTGTAAAGGCCCAAGACCTGAAGGCTTCTCCCGCACCCCTTCCGCC 1517
QY 241 PROGLYGLYASPARGLYARG1YGLYPROGLYGLYMETARG1YGLYARG1YGLYLEU 260
DB 1518 CCGGTGTGATCTGTGCGAGAGTGGCCCTGTGTGATGATCGGGAGAGAGAGTGGCTC 1577
QY 261 METAPARG1YGLYPROGLYGLYMETPHEARG1YGLYARG1YGLYASPLARG1YGLY 280
DB 1578 ATGATCTGTGTGTCTCCCGTGAATGTTCAAGAGGTGGCGGTGTGAGACAGAGGTGG 1637
QY 281 PHEARG1YGLYARG1YMETASPARARG1YGLYPHEGLYGLYGLYARGARG1YGLYPRO 300
DB 1638 TTCGTGTGTGCGGGGCGATGAGCCAGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1697
QY 301 GLYGLYPROPROGLYPROLEUMETGLUG1METGLYGLYARGARG1YGLYARG1YGLY 320
DB 1698 GGGGGGCCCCCTGAGACTTTGATGGAACAGATGGAGAGAAAGAGAGAGAGAGAG 1757
QY 321 PROGLYLYMETASPLYGLYGLYH1SARG1YGLYARGARG1YGLYARG1YGLY 338
DB 1758 CCGGAAAAATGATTAAGCGACAGCCCTCAGAGCGAGAGATCGCCCTTAC 1811

RESULT 2
ID AAS62262/C
ID AAS62262 standard; cDNA; 2176 BP.
XX
XX AAS62262;
XX
XX 14-FEB-2002 (first entry)
XX
XX
DE cDNA sequence #49 encoding novel human secreted protein.
XX
XX
KW Human secreted protein; hyperproliferative disorder; autoimmune disorder;
KW immune deficiency disorder; blood disorder; inflammatory disorder;
KW infectious disorder; gene therapy; antimicrobial; hepatotropic;
KW immunosuppressive; antineumatic; ss.
XX
XX Homo sapiens.
XX
XX MO200177291-A2.
XX
XX 18-OCT-2001.
XX
XX 29-MAR-2001; 2001WO-US010485.
XX
XX 06-APR-2000; 2000US-0195604P.
XX
XX (GENY) GENETICS INST INC.
XX
XX Wong GG, Clark HF, Fectel K, Agostino MJ, Howes SH, Resnick RJ;
XX Gulukota K, Graham JR;
XX
XX MPI; 2002-010900/01.
XX
XX
PT New polynucleotides encoding secreted proteins useful for treating e.g.
PT asthma, HIV and Crohn's disease.
XX
XX
PS Claim 1; Page 100; 391bp; English.
XX
XX
CC The present invention relates to the isolation of novel cDNA sequences
CC which encode human secreted proteins. The cDNA sequences have been
CC derived from a variety of human tissues. The invention also provides a
CC method for producing proteins from these polynucleotide sequences. The
CC proteins are useful for identifying compounds that modulate their
CC activity and production, and the cell is also useful for identifying
CC compounds that modulate expression of the polynucleotide sequences
CC encoding the secreted proteins. The sequences of the invention are useful
CC for treating diseases such as hyperproliferative disorders (e.g. cancer),
CC immune deficiency disorders (e.g. severe combined immunodeficiency
CC (SCID)), autoimmune disorders (e.g. multiple sclerosis), blood disorders

CC (e.g. thrombocytopenia), inflammatory disorders (e.g. arthritis) and
 CC infectious disorders (e.g. hepatitis). The polynucleotide sequences of
 CC the invention are also useful in gene therapy. AAS62214-AAS62838
 CC represent the cDNA sequences of the invention that encode for novel human
 CC secreted proteins

XX
 SQ Sequence 2176 BP; 413 A; 617 C; 563 G; 583 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	2, 01e-60	Length:	2176
Score:	1922.00	Matches:	338
Percent Similarity:	100.004	Conservative:	0
Best Local Similarity:	100.004	Mismatches:	0
Query Match:	100.004	Indels:	0
DB:	6	Gaps:	0

US-10-791-017a-2_COPY_319_656 (1-338) x AAS62262 (1-2176)

QY 1 G1YGLYARGLYGLYMETGLYSEALAGLYGLUARGLYGLYPHEANLYSPROGLYGLY 20
 DB 1207 GAGGAGCGCGGTGAATGGGACGCTGGAGACGAGTGGCTTCAATTAAGCTGTGGA 1148
 QY 21 PromeFASPGLUGLYPROAEPLEUAAPLEUGLYPROPROVALASPROMAPGLUAAPSER 40
 DB 1147 CCCATGGATGAAGACCAAGTCTTGATCTAGGCCCACTGTATCCAGATGAAGACTCT 1088
 QY 41 ASPAENSERALALETYRVALGLINGLYLEUASNAAPSERVALTHIRLEUAAPLEUALA 60
 DB 1087 GACAAACAGTCATTTATGTACAAAGATTAAATGACAGTGTGACTGATGATGTGGCA 1028
 QY 61 AEPHPHEPHELYSEGLYVALVLYVMEETASNLVARGHTRGLYGLNPROMETILE 80
 DB 1027 GACTCTTTTAAGAGTGGGTGTGTTAAGATTAACAAGAACTGGCAACCATGATC 968
 QY 81 HSEILETYRLEUAAPLYSEGLUTHTGLYLYSPROLYGLYASPAATHRVASERTYRGLU 100
 DB 967 CACATCTACCTGGAACAAGAAACAGAAAGCCCAAGAGCATGCCACAGTGTCTTGA 908
 QY 101 AEPHPROTHRALALYVALAVALGLUTTPHESPGLYLYSAPPHESGLYSEY 120
 DB 907 GACCCACCACTGCCAAGGCTGCCGTGGAATGTTGATGGAAAGATTTTCAAGGAGAC 848
 QY 121 LYSLEULYVALSERLEUALAARGLYSLVSPROMETASNSERTMETARGGLYGLYLEU 140
 DB 847 AAACCTTAAGTCTCCCTTGCTCGAAAGAACCTTCATATGAACGATAGCGGTGTCTG 788
 QY 141 PROPROARGGLUGLYARGLYMETPROPROPROLEUARGGLYGLYPROGLYGLY 160
 DB 787 CCAACCCCGTGAAGGAGAGGAGCAACCACTCCGTGAGAGTCCAGAGGCCCAAGGA 728
 QY 161 G1YPROGLYGLYPROMETGLYARGMETGLYLYARGGLYGLYASPARGLYGLYPHEPRO 180
 DB 727 GGTCTGGGGAGCCCATGGGTCCGATGGAGGCGCGTGAAGATAGAGAGGCTTCCT 668
 QY 181 PROARGGLYPROARGGLYSEARGLYASNPPOSEGLYGLYLYVANVALGINHSAAG 200
 DB 667 CCAAGAGGAGCCCGGGGTCTCCGAGGAAACCTCTGAGAGAGAAACGCCGACCA 608
 QY 201 ALAGLYASPTROGLYCYSPROAENPROGLYCYGGLYASNDIANSNPHEALATPARGTHR 220
 DB 607 GCTGGAGACTGGCAAGTGTCCCATCCGGGTGTGGAACCAAGACTTCGCTGGAAGCA 548
 QY 221 GLUCYASNDINCYALYVALAPROLYSPROGLYGLYPHELEUBROPROPROPEPRO 240
 DB 547 GAGTGCACACAGGTAAAGGCCCAAGACCTGAGAGGCTTCCCTCCGACACCTTCCGCC 488
 QY 241 PROGLYGLYASPARGLYARGGLYGLYPROGLYGLYMETARGGLYGLYARGGLYGLYLEU 260
 DB 487 CCGGGTGTGTATGTGTGCAAGAGGTGGCCCTGTGTGCAAGTGGAGAGAGAGGTGGCTTC 428
 QY 261 METASPARGLYGLYPROGLYGLYMETPHEARGGLYGLYARGGLYGLYASPARGLYGLY 280
 DB 427 ATGGATCGTGTGTGTCCCGGTGGAATGTTCAAGAGTGTGGCTGTGGAACAGAGGTGGC 368

QY 281 PHEARGGLYGLYARGLYMETASPARGLYGLYPHEGLYGLYGLYARGARGGLYGLYPRO 300
 DB 367 TTCCTGTGTGTGCGGGGCGATGACCCAGAGTGGCTTGTGTGAGAGAAACAGAGTGGCCCT 308
 QY 301 G1YGLYPROPROGLYPROLEUWETGLUGLYMETGLYGLYLYARGARGGLYGLYARGGLYGLY 320
 DB 307 GGGGGGCCCCCTGGACCTTGTATGAGAACAGATGGAGAAAGAGAGAGGACGCTGGAGGA 248
 QY 321 PROGLYLYMETASPARLYSEGLYGLYHISARGGLYGLYGLYARGARGPARPROTYR 338
 DB 247 CCTGAAAATGATTAAGCGAGCAACCGTCAGAGACGCAAGATGGCCCTTAC 194

RESULT 3

ID AAO50643 standard; cDNA; 2371 BP.

AC AAO50643;
 XX
 DT 25-MAR-2003 (revised)
 DT 26-MAY-1994 (first entry)
 XX
 DE Human Ews gene clone BFLAC5 from foetal brain cDNA library.
 XX
 KW chromosomal translocation; chimeric; chimaeric; Ewing sarcoma; Ews gene;
 KW malignant melanoma; hum-Flt-1;
 KW primitive peripherai neuroectodermal tumour; human chromosome 11;
 KW human chromosome 22; ss.
 XX
 OS Homo sapiens.

XX Key Location/Qualifiers

PH CDS 25..1995
 FT /*tag= a
 FT /product= "EWS protein"
 FT /transl_except= pos:1729..1731; aa:Val
 FT polyA_signal 2143..2148
 FT polyA_signal 2331..2336
 FT /*tag= c
 PN MO9323549-A2.

XX 25-NOV-1993.
 XX 19-MAY-1993; 93WO-FR000494.
 XX PF
 XX PR 20-MAY-1992; 92FR-00006123.
 XX

XX (CNRS) CNRS CENT NAT RECH SCI.

XX Aurias A, Delattre O, Desmaza C, Melot T, Peter M, Plougastel B;
 PI Thomas G, Zucman J;
 XX

DR WPI; 1993-386580/48.
 DR P-PSDB; AAR44555.
 XX

PT New nucleic acid of EWS gene and its hybrid(s) - contg. gene sequence
 PT involved in chromosomal trans-location, also derived mRNA, probes, fusion
 PT proteins etc., for diagnosis and treatment of Ewing sarcoma and melanoma.
 XX
 PS Disclosure; Fig 6; 123pp; French.

XX The probes 22R3 and 22R12 were used to screen a human foetal brain cDNA
 CC library (Stratagene cat. # 936206). The clone BFLAC5 was identified and
 CC sequenced. It represents the entire coding region and 3'-UTR of the Ews
 CC gene. (Updated on 25-MAR-2003 to correct PN field.)
 XX

SQ Sequence 2371 BP; 639 A; 587 C; 659 G; 486 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	2, 13e-60	Length:	2371
Score:	1922.00 <td>Matches:</td> <td>338</td>	Matches:	338

Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 2 Gaps: 0

US-10-791-017a-2_copy_319_656 (1-338) x AAQ50643 (1-2371)

QY 1 G1YGLVARG1YGLYMETG1YSEALAG1YGLVARG1YGLYVHEANLYSEPROG1YGLY 20
 DB 979 GGAGGAGCGGTGGAATGGGACGGCTGAGAGGAGGAGTGGCTTCAATTAAGCCCTGGTGA 1038
 QY 21 PROMETAPGLUG1YPROAPLEUAPLEUG1YPROPROVAL1APPROAGG1UAPSER 40
 DB 1039 CCCATGATGAAGAACCAAGATCTTGATCTAGGCCCTCTTGAATCCAGATGAAGACTCT 1098
 QY 41 AAPANSEAL1A1ETRYVAL1G1NG1YLEUAPNAPSERVAL1TH1EUAAPLEU1A 60
 DB 1099 GACAAACAGTCAATTTATGTATCAAGATTAATGACAGTGTGACTTATGATGATCTGGCA 1158
 QY 61 AAPHPEHEL1YSG1YVAG1YVAL1YVMEFAS1YVARG1YGL1N1PROMET1LE 80
 DB 1159 GACTTCTTAAAGCAGTGTGGGTGTGTTAAAGTAAAGAACTGGGCAACCATGATC 1218
 QY 81 H1E11ETRYLEUAP1YGL1YTH1YLYSEPRO1YGL1YAP1A1TH1V1SE1YTRG1U 100
 DB 1219 CACATCTACCTGACAAAGAAACAGAAACCCAAAGCCGATGCCACAGTCTATGA 1278
 QY 101 AAPPROBOTH1A1ALY1A1A1A1A1A1G1UT1PHEAP1Y1YVAP1PHEG1NG1YSE 120
 DB 1279 GACCCACCCACCTGCAAGGCTGGCCGTGGAATGTTGATGGAAAGATTTTCAAGGGAGC 1338
 QY 121 LY1E1U1Y1V1A1SE1LEU1A1A1Y1YV1SEPRO1YGL1YVHEANLYSE1YGL1YLEU 140
 DB 1339 AAACCTTAAGTCTCCCTTGCTGCGAAGACCTCTCAATGAACAGTATGCGGGGTGCTG 1398
 QY 141 PROPROARG1YGL1YARG1YME1YPRO1YPROLEUARG1YGL1YPROG1YGL1Y 160
 DB 1399 CCACCCCGTAGAGGACAGGACATCCACACACTCCGTGAGAGTCCAGAGGCGCCACGA 1458
 QY 161 G1Y1P1R1G1YGL1Y1P1R1M1E1Y1Y1V1M1E1Y1Y1V1A1R1G1YGL1Y1P1H1E1P1R1O 180
 DB 1459 GGTCTCTGGGAGACCCATGAGGTGCGATGGAGGCGGTGAGAGATGAGAGAGGCTTCCCT 1518
 QY 181 PROARG1YPROARG1YSE1YARG1YVAP1YVAP1YGL1YV1A1N1V1G1N1H1A1R1G 200
 DB 1519 CCAGAGAGACCCCGGGGTCTCCCGAGGAAACCTCTGAGAGAGAAACGTCACAGACCGA 1578
 QY 201 A1AG1YAP1TP1G1YCY1P1R1O1YAP1P1R1O1YCY1Y1A1N1G1N1H1E1N1P1H1E1A1T1P1R1G1T1H 220
 DB 1579 GCTGGAGAC1YGGCAGTCTCCCAATCCGGGTTGTGGAACCAAGAACTTCCGCTGGAGAA 1638
 QY 221 G1UCY1A1N1G1Y1V1A1A1P1R1O1Y1P1R1O1Y1P1H1E1U1P1R1O1P1H1E1P1R1O 240
 DB 1639 GAGTGCACCAAGTGTAAAGGCCCAACCTGAAAGGCTCTCCCGCCACCCCTTCCGCCCC 1698
 QY 241 PROG1YGL1YAP1YARG1YVARG1YGL1YVHEANLYSE1YVARG1YGL1YLEU 260
 DB 1699 CCGGAGTGTATCTGAGGAGAGG1YGGCCCTG1YGGCA1YGGGAGGAGAAAGG1YGGCTC 1758
 QY 261 ME1E1AP1R1G1YGL1Y1P1R1O1Y1Y1M1E1Y1VHEANLYSE1YVARG1YGL1Y 280
 DB 1759 ATGATGTGTGTGTCCCGG1YGAATTTTCAAGAGTGGCCGTGTGAGACAGAGTGGC 1818
 DB 1819 TTCCTGTGTGTGCGCGGACATGACCGAGTGGCTTTGTGTGAGAGAAACAGAGTGGCCCT 1878
 QY 301 G1Y1Y1P1R1O1Y1P1R1O1Y1P1R1O1Y1P1R1O1Y1P1R1O1Y1P1R1O1Y1P1R1O1Y1 320
 DB 1879 GGGGGGGCCCCCTGGACCTTGTATGAAACAGATGGAGAGAAAGAGAGAGACG1YGGAGGA 1938
 QY 321 PROG1YVHEANLYSE1YGL1YH1E1A1R1G1N1G1Y1VARG1YVAP1YPRO1Y 338

DB 1939 CCGGAAAAATGATTAAGGCGAGCACGCTCAGAGCGGAGAGATCGCCCTAC 1992
 RESULT 4
 ADQ86032
 ID ADQ86032 standard; cDNA; 2372 BP.
 AC ADQ86032;
 DT 07-OCT-2004 (first entry)
 DE Human tumour-associated antigenic target (TAT) cDNA sequence #2904.
 XX human; tumour-associated antigenic target; TAT; cytostatic; gene therapy;
 KW cancer; cell proliferative disorder; gene; ss.
 OS Homo sapiens.
 PN WO2004060270-A2.
 XX
 PD 22-JUL-2004.
 XX
 PF 15-OCT-2003; 2003WO-US029126.
 XX
 PR 18-OCT-2002; 2002US-0418988P.
 XX
 PA (GENTH) GENENTECH INC.
 PA (WUTD/) WU T D.
 PA (ZHOU/) ZHOU Y.
 XX
 PI Wu TD, Zhou Y;
 XX
 DR WPI; 2004-534300/51.
 XX
 PT New nucleic acid molecule and encoded polypeptide, for diagnosing,
 PT preventing or treating cell proliferative disorders such as cancer.
 PS
 XX
 PS Claim 1; SEQ ID NO 2904; 5504bp; English.
 XX
 CC The present invention describes an isolated tumour-associated antigenic
 CC target (TAT) nucleic acid comprising: (a) any of 4622 nucleotide
 CC sequences (see SEQ ID NO:1 to 4622); (b) the full-length coding region of
 CC (a); (c) the complement of (a) or (b); (d) a sequence that has 80%
 CC sequence identity to (a)-(c); or (e) a sequence that hybridizes to (a)-
 CC (c). Also described: (1) an expression vector comprising the above
 CC nucleic acid; (2) a host cell comprising the above expression vector; (3)
 CC a process for producing a polypeptide; (4) an isolated polypeptide
 CC comprising: (a) an amino acid sequence encoded by any of the above
 CC nucleotide sequences; (b) an amino acid sequence encoded by the full-
 CC length coding region of the above nucleotide sequences; or (c) a sequence
 CC having at least 80% identical to (a) or (b); (5) a chimeric polypeptide
 CC comprising the above polypeptide fused to a heterologous polypeptide; (6)
 CC an isolated antibody that binds to the above polypeptide; (7) a process
 CC for producing the antibody; (8) an isolated oligopeptide that binds to
 CC the above polypeptide; (9) a tumour-associated antigenic target (TAT)
 CC binding organic molecule that binds to the above polypeptide; (10) a
 CC composition of matter comprising the above (chimeric) polypeptide;
 CC antibody, oligopeptide or TAT binding organic molecule, in combination
 CC with a carrier; (11) an article of manufacture comprising a container and
 CC the composition of matter contained within the container; (12) methods of
 CC inhibiting the growth of a cell that expresses the above protein, where
 CC the growth of the cell is at least in part dependent upon a growth
 CC potentiating effect of the above protein; (13) a method of
 CC therapeutically treating a mammal having a cancerous tumour comprising
 CC cells that express the above protein; (14) a method of determining the
 CC presence of a protein in a sample suspected of containing the protein
 CC described above; (15) methods of diagnosing the presence of a tumour in a
 CC mammal; (16) a method for treating or preventing a cell proliferative
 CC disorder associated with increased expression or activity of the above
 CC protein; and (17) a method of binding an antibody, oligopeptide or
 CC organic molecule to a cell that expresses the protein described above.
 CC The TAT sequences have cytostatic activities, and can be used in gene
 CC therapy. The composition and methods are useful for diagnosing,
 CC preventing or treating cancer. The composition is also used for preparing

CC a medicament for the therapeutic treatment or diagnostic detection of a
 CC cell proliferative disorder or cancer. The present sequence represents a
 CC human TAT cDNA sequence from the present invention.

XX Sequence 2372 BP; 628 A; 589 C; 668 G; 487 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	2,13e-60	Length:	2372
Score:	1922.00	Matches:	338
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	13	Gaps:	0

US-10-791-017a-2_COPY_319_656 (1-338) x ADQ86032 (1-2372)

QY 1 G1YGYARGLYGLYMETGLYSEALAGLYGLUARGLYGLYPHEANLYSPROGLYGLY 20
 DB 998 GGAGAGCGCGTGAATGGGCAAGCGCTGGAGAGAGAGTGGCTTCAATTAAGCGTGGGA 1057
 QY 21 PROMETAPROGLYGLYPROAPLEUAPLEUGLYPROPROVALAPPROAPGLUAPSER 40
 DB 1058 CCATGGATGAGGACCAAGATCTTGAATCTAGGCCCACTGTAGATCCAGATGAAGACTCT 1117
 QY 41 ASPANSEALATLETYRVALGLINGLYLEUASAPSERVALPHLEUASAPLEUALA 60
 DB 1118 GACACACAGTCATTTATGTAACAAGATTAATGACGTGTAGCTTGAATGATCTGGCA 1177
 QY 61 ASPHAPHELYSGLYCYSGLYVALLYAMECANLYSARGLYTHRGLYGLINPROMETILE 80
 DB 1178 GACTTCTTTAAGCAGTGTGGGTTGTTAAGATGAACAAGAACTGGCAACCATGATC 1237
 QY 81 HSEILETYRLEUAPLYGLYTHRGLYLYSPROLYSGLYSAPALATHRVALSERTYGLU 100
 DB 1238 CACATCTACTGACCAAGAAACAGGAAAGCCCAAGCCATGACCAAGTCTCTATGAA 1297
 QY 101 ASPPROPTHRALALYSAALALVALGLUTRPHAPGLYLYSAPPHENGLYGLYSER 120
 DB 1298 GACCCACCCACCTGCAAGGCTGCGGTGGAATGGTTTATGGAAGATTTTCAAGGAGC 1357
 QY 121 YLEULYVALSERLEUALAARGLYLYSPROLYSPROLYSPROLYSPROLYSPROLYSPROLY 140
 DB 1358 AAATTTAAAGTCTCTCTGCTCGGAAGAACCTTCAATGAACAGTANCGGGGTGCTG 1417
 QY 141 PROPROARGLYGLYGLYMETPROPROPROLEUARGLYGLYGLYGLYGLYGLYGLY 160
 DB 1418 CCACCCCGTGAAGGAGAGGATGCGCACCACTCGTGAAGGTCCAGAGAGGCCCGCA 1477
 QY 161 GLYPROGLYGLYSPROLYSPROLYSPROLYSPROLYSPROLYSPROLYSPROLYSPROLY 180
 DB 1478 GGTCTCGGGGAGCCCATGGATCGCATGGAGGCGGTGGAAGATGAAGAGGCTTCCCT 1537
 QY 181 PROARGLYPROARGLYSPROLYSPROLYSPROLYSPROLYSPROLYSPROLYSPROLY 200
 DB 1538 CCAAGAGAGCCCGGGGTTCCCAAGGAGAACCTCTGGAAGAGAGAACCTCCAGACCGA 1597
 QY 201 ALAGLYAPPTPGLYGLYSPROLYSPROLYSPROLYSPROLYSPROLYSPROLYSPROLY 220
 DB 1598 GCTGAGAGACTGCGAGTGTCCCAATCCGGGTTGTGGAACAGAACTCCGCTGGAAGACA 1657
 QY 221 GLUCYASAGNLYGLYSPROLYSPROLYSPROLYSPROLYSPROLYSPROLYSPROLYSPROLY 240
 DB 1658 GAGTGCACCAAGTGAAGGCGCCCAAGGAGGCTTCCCGGCAACCTTCCGCCCC 1717
 QY 241 PROGLYGLYSPROLYSPROLYSPROLYSPROLYSPROLYSPROLYSPROLYSPROLYSPROLY 260
 DB 1718 CCGGTGTGTATCTGTGCAAGGTGGCTCTGTGTGCAAGGTGGGAGAGAGAGGTGCTC 1777
 QY 261 METAPAPARGLYGLYGLYGLYGLYGLYGLYGLYGLYGLYGLYGLYGLYGLYGLYGLY 280
 DB 1778 ATGATGATGT 1837
 QY 281 PHEARGLYGLYGLYGLYGLYGLYGLYGLYGLYGLYGLYGLYGLYGLYGLYGLYGLYGLY 300

DB 1838 TTCCGTGTGTGCGCGGAGCAGACCAAGTGTGCTTGTGTGAGAGAGACAGAGTGGCCT 1897
 QY 301 G1YGYARGLYGLYMETGLYSEALAGLYGLUARGLYGLYPHEANLYSPROGLYGLY 320
 DB 1898 GGGGGGCGCCCTGAGCTTGTGATGGAACAGATGAGGAGAGAGAGAGAGAGAGAGAGAG 1957
 QY 321 PROGLYLYMETAPROGLYGLYGLYGLYGLYGLYGLYGLYGLYGLYGLYGLYGLYGLY 338
 DB 1958 CCGGAAAATGATTAAGGCGAGACCGTCAGAGGCGAGAGATCGGCTTAC 2011
 RESULT 5
 AAS70647
 ID AAS70647 standard; cDNA; 2388 BP.
 AC AAS70647;
 DT 13-FEB-2002 (first entry)
 XX
 DE DNA encoding novel human diagnostic protein #6451.
 XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
 XX Homo sapiens.
 OS
 FN W0200175067-A2.
 PD 11-OCT-2001.
 XX
 PF 30-MAR-2001; 2001WO-US008631.
 XX
 PR 31-MAR-2000; 2000US-00540217.
 PR 23-AUG-2000; 2000US-00649167.
 XX
 PA (HYSE-) HYSEQ INC.
 PI Dmanac RT, Liu C, Tang YT;
 XX
 DR WPI; 2001-639362/73.
 XX
 DR P-PSDB; ABG06460.
 XX
 PT New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity.
 XX
 PS Claim 1; SEQ ID NO 6451; 103pp; English.
 XX
 CC The invention relates to isolated polynucleotide (I) and polypeptide (II)
 CC sequences. (I) is useful as hybridisation probes, polymerase chain
 CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
 CC in recombinant production of (II). The polynucleotides are also used
 CC in diagnostics as expressed sequence tags for identifying expressed
 CC genes. (I) is useful in gene therapy techniques to restore normal
 CC activity of (II) or to treat disease states involving (II). (II) is
 CC useful for generating antibodies against it, detecting or quantitating a
 CC polypeptide in tissue, as molecular weight markers and as a food
 CC supplement. (II) and its binding partners are useful in medical imaging
 CC of sites expressing (II). (I) and (II) are useful for treating disorders
 CC involving aberrant protein expression or biological activity. The
 CC polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic
 CC coding sequences of the invention. Note: The sequence data for this
 CC patent did not appear in the printed specification, but was obtained in
 CC electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 SO Sequence 2388 BP; 628 A; 599 C; 668 G; 493 T; 0 U; 0 Other;

Alignment Scores:

Prod. No.:	2,14e-60	Length:	2388
Score:	1922.00	Matches:	338
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	5	Gaps:	0

US-10-791-017a-2_copy_319_656 (1-338) x AAS70647 (1-2388)

```

QY 1 G1YGLYARGLYGLYMETGLYSERLALGLYGLUARGLYGLYPhEaNLyPProGLYGLY 20
DB 1015 GGAGGACGCGGTGAATGGGCGAGCGCTGGAGAGCGAGTGTCAATAGCCCTGGGGA 1074

QY 21 ProMetAspGluGlyProAspLeuAspLeuGlyProProValAspProAspGluAspSer 40
DB 1075 CCAATGATAGAGACCAAGATCTTGATCTAGCCCTCTTAAGTCCAGATGAGACTCT 1134

QY 41 AspAsnSerAlaIleTyrValGlnGlyLeuAsnAspSerValThrLeuAspAspLeuAla 60
DB 1135 GACAAACAGTCAATTATATGACAAGATTAATGACAGTGTGACTCTAGATGATCTGGCA 1194

QY 61 AspPhePheLysGlnCysGlyValLysMetAsnLysArgThrGlyGlnProMetIle 80
DB 1195 GACTTCTTAAGCAGTGTGGGTTGTTAAGATGAAACAAGAACTGGGCAACCATGATC 1254

QY 81 HisIleTyrLeuAspLysGluThrGlyLysProLysGlyAspAlaThrValSerTyrGlu 100
DB 1255 CACATCTACTGACCAAGAAAGCAAGAAAGCCCAAGACCATGCCACAGTCTCTATGAA 1314

QY 101 AspProThrAlaLysAlaLysAlaValGluTyrPheAspGlyLysAspPheGlnGlySer 120
DB 1315 GACCCACCCGACGCAAGGAGCGTCCGCTGGAATGTTTATGGAAGAAAGATTTTCAAGGAGC 1374

QY 121 LysLeuLysValSerLeuAlaArgLysLysProPheAsnSerMetAspGlyLysLeu 140
DB 1375 AAACCTTAAGTCTCCCTTCTCGGAGAGAACCTTCATGAAAGTATGCGGGTGGTCTG 1434

QY 141 ProProArgGluGlyArgGlyMetProProProLeuAspGlyGlyProGlyGlyProGly 160
DB 1435 CCACCCCGTAGAGGACAGGAGCATGCAACCACTCGTAGAGGTCCAGAGGCGCCAGGA 1494

QY 161 G1YProGlyLysProMetGlyArgMetGlyLysArgGlyLysLysAspArgGlyLysPhePro 180
DB 1495 GGTCTGGGGGAGCCCATGGGTCGATGGAGGCGCGTGGAGAGATGAGAGAGGCTTCCT 1554

QY 181 ProArgGlyLysProArgGlySerArgGlyAsnProSerGlyGlyLysAsnValGlnHisArg 200
DB 1555 CCAAGAGGAGCCCGGGGCTTCCCGAGGAGAACCCCTCTGAGAGAGAAAGCTCCAGACCGA 1614

QY 201 AlaGlyAspTyrGlnCysProAsnProGlyCysGlyAsnGlnAsnPheAlaTyrArgThr 220
DB 1615 GCTGGAGACTGGAGGTGTCCCAATCCGGGTGTGGAAACAAGAACTTCGCTGGAGAGCA 1674

QY 221 GlnCysAsnGlnCysLysAlaProLysProGlyLysPheLeuProProPheProPhePro 240
DB 1675 GAGTGCAACAGGTAGAGGCCCAAGACCTGGAAGGCTTCTCCGCAACCTTTCCGCCCC 1734

QY 241 ProGlyLysLysAspArgGlyArgGlyLysProGlyLysMetArgGlyLysArgGlyLysLeu 260
DB 1735 CCGGTGTGTATCTGTGCAAGAGGTGGCTGTGGTGCATGCGGGAGAGAAAGGTGGCTC 1794

QY 261 MetAspArgGlyLysLysProGlyLysMetPheArgGlyLysLysArgGlyLysLys 280
DB 1795 ATGATATGTGTGTCCCGGTGGAGATTTCAAGAGTGGCGGTGTGAGACAAGAGTGGC 1854

QY 281 PheArgGlyLysLysArgGlyMetAspArgGlyLysPheGlyLysLysArgGlyLysPro 300
DB 1855 TTTCGTGTGTGGCCGGGCAATGACCAAGAGTGGCTTTGTGTGAGAGAAACAAGTGGCCCT 1914

QY 301 G1YGLYProProGlyLysProLeuMetGlnGlnMetGlyLysLysArgGlyLysArgGlyLys 320
DB 1915 GGGGGGGCCCTTGACCTTGTATGAGAACATGTGGAGAGAAAGAGAGAGACGTGGAGGA 1974

```

```

QY 321 ProGlyLysMetAspLysGlyLysHisArgGlnGlnArgArgAspArgProTyr 338
DB 1975 CTTGGAAAATGATTAAGGCGAGCACCTCTCAGAGCGCAGAGATCGCCCTTAC 2028

RESULT 6
ID ABR84628
ABR84628 standard; cDNA; 2390 BP.
AC ABR84628;
XX
XX
XX 14-AUG-2002 (first entry)
XX
XX
XX Human cDNA differentially expressed in granulocytic cells #1199.
DB
XX
XX Human; ss; granulocytic cell; DNA chip; bacterial infection;
XX viral infection; parasitic infection; protozoal infection;
XX fungal infection; sterile inflammatory disease; psoriasis;
XX rheumatoid arthritis; glomerulonephritis; asthma; thrombosis;
XX cardiac reperfusion injury; renal reperfusion injury; ARDS;
XX adult respiratory distress syndrome; inflammatory bowel disease;
XX Crohn's disease; ulcerative colitis; periodontal disease;
XX granulocyte activation; chronic inflammation; allergy.
XX
XX Homo sapiens.
XX
XX WO200228999-A2.
XX
XX 11-APR-2002.
XX
XX 03-OCT-2001; 2001WO-US030821.
XX
XX 03-OCT-2000; 2000US-0237189P.
XX
XX (GENE-) GENE LOGIC INC.
XX
XX Beazer-Barclay Y, Weissman SM, Yamaga S, Vockley J;
XX
XX WPI; 2002-435328/46.
XX
XX Detecting granulocyte activation by detecting differential expression of
XX genes associated with granulocyte activation, which serves as diagnostic
XX markers that is useful for monitoring disease states and drug toxicity.
XX
XX Claim 1; SEQ ID NO 1199; 114pp; English.
XX
XX The invention relates to detecting (M1) granulocyte (GC) activation
XX (GCA), by detecting the level of expression of gene(s) (Gs) identified by
XX DNA chip analysis as given in the specification, and comparing the
XX expression level to an expression level in an unactivated GC, where
XX differential expression of Gs is indicative of GCA. Also included are
XX modulating (M2) GA by contacting GC with an agent that alters the
XX expression of at least one gene in Gs; (2) screening (M3) for an agent
XX capable of modulating GCA or an inflammation (especially chronic) in a
XX tissue, an allergic response in a subject, exposure of a subject to a
XX pathogen or sterile inflammatory disease using the gene expression
XX profile; (3) detecting (M4) an inflammation (especially chronic) in a
XX tissue, an allergic response in a subject, exposure of a subject to a
XX pathogen or sterile inflammatory disease, by detecting the level of
XX expression in a sample of the tissue of gene(s) from Gs, where the level
XX of expression of the gene is indicative of inflammation; (4) treating
XX (M5) an inflammation (especially chronic) or in a tissue, an allergic
XX response in a subject, exposure of a subject to a pathogen or sterile
XX inflammatory disease, by contacting a tissue having inflammation with an
XX agent that modulates the expression of gene(s) from Gs in the tissue. M1
XX is useful for detecting GCA; M2 is useful for modulating GA; M3 is useful
XX for screening an agent capable of modulating GCA preferably in an
XX inflammation in a tissue; M4 is useful for detecting an inflammation
XX (especially chronic) in a tissue, an allergic response in a subject,
XX exposure of a subject to a pathogen or sterile inflammatory disease (e.g.
XX psoriasis, rheumatoid arthritis, glomerulonephritis, asthma, thrombosis,
XX cardiac reperfusion injury, renal reperfusion injury, ARDS, adult
XX respiratory, distress syndrome, inflammatory bowel disease, Crohn's

```


CC prostate cancer, ABK64106-ABK64860 represent human benign prostatic
CC hyperplasia gene sequences of the invention
XX
SQ Sequence 2390 BP, 645 A, 589 C, 668 G, 488 T, 0 U, 0 Other;

Alignment Scores:

Pred. No.:	2,14e-60	Length:	2390
Score:	1922.00	Matches:	338
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	6	Gaps:	0

US-10-791-017a-2_COPY_319_656 (1-338) x ABK64822 (1-2390)

QY 1 GYGLYARGLYGLYMETGLYSERAGLYGLUARGLYGLYPHEANLYSPROGLYGLY 20
DB 998 GGAGGACGCGGTGAATGGCGACGCTGAGACGAGGTGGCTTCAATAAGCCGTGGGA 1057
QY 21 PROMETASPGLUGLYPROAEPLEUAPLEUGLYPROBVALAPPROAPBGLUAPSER 40
DB 1058 CCCATGGATAGAGACCAATCTTGATCTAGCCCTCTGTAGATCCAGATGAGACTCT 1117
QY 41 ASPASERLALLETYRVALGNGLYLEUASAPSERVALTHLEUASAPLEUALA 60
DB 1118 GACACAGTCGCAATTATATGACAGATTAATGACAGTGTGACTTATGATGATCTGGCA 1177
QY 61 ASPHEPHELYGLNCGLYVALVALLYMECAENLYAARGTHRGLYNPROMETILE 80
DB 1178 GACTTCTTTAAGCAGTGTGGGTGTTTAAAGATGAACTGGGCAACCCATGATC 1237
QY 81 HIAILETYRLEUASPLYSGLUUTHRGLYLYSPROLYSGLYASPLAATHVALSERTYRGLU 100
DB 1238 CACATCTACTGACCAAGAAACAGAAAGCCCAAGCCAGATGACAGTGTCTATGAA 1297
QY 101 ASPPROBOTHRALALYSAALAAVALGLUTRPHASPLGLYLYASAPHEGLNGLYSER 120
DB 1298 GACCCACCCACTGCGCAAGGCTGCGGTGGAATGGTTTATGGAAAGATTTTCAAGGAGC 1357
QY 121 LYLEULYVALSERLEUALAARGLYLYSPROBROMECANBSEMETATRGLYLYLEU 140
DB 1358 AAACCTTAAGTCTCCCTGCTCGGAGAAACCTTCAATGAACTATGCGGGGTGCTG 1417
QY 141 PROBPARGLUGLYARGLYMETPROBPROBLEUARGLYGLYPROGLYGLYPROGLY 160
DB 1418 CCAACCCGTAAGGACAGAGGATGCAACACTCCGTGAGGTCCAGAGGCGCAGGA 1477
QY 161 GLYPROGLYGLYPROMETGLYARGMETGLYGLYARGLYGLYASPARGLYGLYPHEPRO 180
DB 1478 GGTCTGGGGGACCCATGGGTGCGATGGAGGCGGTGAGAGATGAGAGAGGCTTCCCT 1537
QY 181 PROARGLYPROARGLYSERARGLYASBPROSERGLYGLYASNVVALGHNHIAARG 200
DB 1538 CCAAGAGAGACCCCGGGGTTCGCCGAGGAAACCCCTCTGAGAGAGGAAACGTCACGCA 1597
QY 201 ALGLYASPTRPGINCYSPROANPROGLYCYGSLYASNGINANPHEALATPARGTHR 220
DB 1598 GCTGGACACTGGCAGTGTCCCAATCCGGGTGTGGAACAAGAACTTGCGCTGAGACA 1657
QY 221 GLUCYASANGINCYLYSALAPROLYSPROGLYGLYLYPHELEUPROPROPHETPRO 240
DB 1658 GAGTGCAACCACTGTAAGGCCCCCAAGCCCTGTAAGGCTTCCCGCCACCCCTTCCGCC 1717
QY 241 PROGLYLYASPARGLYARGLYGLYPROGLYGLYLYMETARGLYGLYARGLYGLYLYLEU 260
DB 1718 CCGGTGTGTATCTGTGACAGAGGTGCGCTGTGTGCAATGCGGGAGAGAAAGGTGCGCTC 1777
QY 261 METASPARGLYGLYPROGLYGLYMETPHEARGLYGLYARGGLYGLYASPARGLYGLY 280
DB 1778 ATGATGTGTGTGTCGGGTGAGATGTCAGAGGTGCGGTGTGAGACAGAGGTGGC 1837
QY 281 PHEARGLYGLYARGGLYMETASPARGLYGLYLYPHEGLYGLYGLYARGARGGLYGLYPRO 300

DB 1838 TTCCTGTGTGCGGGGACATGACCGAGGTGCTTTGTGTGAGAAAGACAGGTGCCCT 1897
QY 301 GYGLYPROBPROGLYPROLEUMETGLUINMETGLYGLYARGARGLYGLYARGGLYGLY 320
DB 1898 GGGGGGCCCCCTGGACCTTGTGATGAAACAGATGGGAGAAAGAGAGAGACCTGAGGA 1957
QY 321 PROGLYLYMECAEPYASGLYGLUHSARGGLYGLUARGARGSPARPROTYR 338
DB 1958 CTTGAAAAATGATTAAGCGACGACCTTCAGAGACGAGATCGGCCCTTAC 2011

RESULT 8

ABN97274 standard; DNA; 2390 BP.

ABN97274;

13-AUG-2002 (first entry)

Gene #3772 used to diagnose liver cancer.

KW Gene; liver cancer; de; hepatocellular carcinoma; hepatotropic;
KW metastatic liver tumor; cytostatic; expression profile; disease state;
KW disease progression; drug toxicity; drug efficacy; drug metabolism.

OS Homo sapiens.

PN WO200229103-A2.

PD 11-APR-2002.

PF 02-OCT-2001; 2001WO-US030589.

PR 02-OCT-2000; 2000US-0237054P.

PA (GENE-) GENE LOGIC INC.

PI Horne D, Alvares C, Peres-Da-Silva S, Vockley JG;

DR WPI; 2002-426119/45.

PT Diagnosing and detecting the progression of liver cancer, hepatocellular
PT carcinoma or metastatic liver tumor in a patient, involves detecting the
PT level of expression of two or more genes in a liver tissue sample.

PS Claim 1; SEQ ID NO 3772; 298bp; English.

XX The invention relates to a novel method for diagnosing and detecting the
XX progression of liver cancer, hepatocellular carcinoma or metastatic liver
XX tumor in a patient, and differentiating metastatic liver cancer from
XX hepatocellular carcinoma in a patient, involving detecting the level of
XX expression of two or more genes represented in ABN93503-ABN97455 in a
XX tissue sample. The method of the invention has hepatotropic, and
XX cytostatic activity. The method is useful for diagnosing and detecting
XX the progression of liver cancer, hepatocellular carcinoma and metastatic
XX liver carcinoma in a patient. The method is useful for identifying
XX expression profiles which serve as useful diagnostic markers as well as
XX markers that can be used to monitor disease states, disease progression,
XX drug toxicity, drug efficacy and drug metabolism. Note: The sequence data
XX for this patent did not form part of the printed specification, but was
XX obtained in electronic format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences

SQ Sequence 2390 BP, 645 A, 589 C, 668 G, 488 T, 0 U, 0 Other;

Alignment Scores:

Pred. No.:	2,14e-60	Length:	2390
Score:	1922.00	Matches:	338
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	6	Gaps:	0

US-10-791-017a-2_COPY_319_656 (1-338) x ABN97274 (1-2390)

```

QY      1  G|YGLYARGLYGLYMETGLYSER|AGLYGLYUARGLYGLY|PHEANLYSPROGLYGLY 20
DB      998  GGAGGAGCGCGGTGAATGGCGACGGCTGGAGAGGAGAGTGCTTCATTAAGCCCTGGTGA 1057
QY      21  P|MEC|A|SP|GLU|GLY|P|RO|A|P|LEU|A|P|LEU|GLY|P|RO|P|RO|VAL|A|P|RO|A|P|GLY|A|P|SER 40
DB      1058  CCCATGGATGAAGACCAAGATCTTGATCTAGGCCCTCTCTAGATCCAGATGAAGATCTT 1117
QY      41  A|P|A|N|S|E|A|L|A|I|E|Y|V|A|L|G|I|N|G|L|Y|L|E|U|A|N|A|P|S|E|V|A|L|T|H|L|E|U|A|P|A|P|L|E|U|A| 60
DB      1118  GACACACGTCGCAATTTATGTATGACAAGAGATTAAATGACAGTGAGCTTACATGATGATCTGCA 1177
QY      61  A|P|P|H|E|L|Y|S|E|G|I|N|C|S|E|G|I|V|A|L|Y|L|E|U|E|N|L|Y|A|P|P|H|E|L|Y|S|E|G|I|N|C|S|E|G|I| 80
DB      1178  GACTCTCTTAAAGCAGTGTGGGTGTTTAAAGTAAAGAACCTGGGCAACCATATATC 1237
QY      81  H|E|I|E|T|Y|L|E|U|A|P|L|Y|S|E|G|I|U|T|H|G|L|Y|L|E|U|P|RO|L|Y|S|E|G|I|V|A|P|A|L|T|H|V|A|L|S|E|T|Y| 100
DB      1238  CACATCTACCTGACCAAGAAACAGAAACGAAAGCCCAAGGCGATGCCACAGTGTCTTATGAA 1297
QY      101  A|P|P|RO|P|RO|T|H|A|L|A|U|Y|A|A|A|A|V|A|L|G|U|T|P|H|E|A|P|G|L|Y|Y|A|P|P|H|E|G|I|N|G|L|Y|S|E| 120
DB      1298  GACCCACCACTGCGCAAGGCTGCGGTGGAATGGTTTATGGGAAAGATTTTCAAGGAGGC 1357
QY      121  L|Y|L|E|U|L|Y|S|A|L|S|E|R|L|E|U|A|A|A|G|L|Y|L|E|U|P|RO|P|RO|E|N|S|E|R|E|T|A|G|L|Y|L|E|U 140
DB      1358  AAACCTTAAGATCTCCCTGCTGGAGAACCTTCCAATGAACAGTATGCGGGGTGCTCTG 1417
QY      141  P|RO|P|RO|A|RG|L|U|G|L|Y|A|RG|L|Y|ME|P|RO|P|RO|L|E|U|A|RG|L|Y|GLY|P|RO|GLY| 160
DB      1418  CCACCCCGTGAAGGCGAGAGGCAATGCCACACACTCCGTGAGAGTCCAGAGAGGCCCAAGA 1477
QY      161  G|Y|P|RO|GLY|GLY|P|RO|ME|T|G|L|Y|A|RG|L|Y|GLY|A|RG|L|Y|GLY|A|P|A|RG|L|Y|GLY|P|H|E|P|RO 180
DB      1478  GGTCTGGGGGAGCCCATGGGTGCGATGAGGAGCGCTGGAGAGATAGAGAGGCTTCCCT 1537
QY      181  P|RO|A|RG|L|Y|P|RO|A|RG|L|Y|S|E|A|RG|L|Y|A|N|P|RO|S|E|T|Y|GLY|V|A|N|V|A|L|G|I|N|H|A|RG 200
DB      1538  CCAAGAGAGACCCCGGGGTTCCTCCAGAGGAAACCCCTCTGAGAGAGAAACGCTCCAGACCGA 1597
QY      201  A|A|G|L|Y|A|P|T|P|G|I|N|C|S|P|RO|A|N|P|RO|GLY|GLY|V|A|N|G|I|N|A|N|P|H|E|A|L|T|P|A|RG|T|H| 220
DB      1598  GCTGGAGACTGGCAGATGATCCCATCCGGGTTGTGGAACCAAGATCTTCCCTGGAGACA 1657
QY      221  G|U|C|Y|A|N|G|I|N|C|Y|L|Y|A|L|A|P|RO|L|Y|P|RO|GLY|GLY|P|H|E|U|P|RO|P|RO|P|H|E|P|RO 240
DB      1658  GAGTGCAACCAAGTGAAGGCCCCAAAGCCTGAAGGCTTCCCTCCGCAACCTTCCGCC 1717
QY      241  P|RO|GLY|GLY|A|P|A|RG|L|Y|A|RG|L|Y|GLY|P|RO|GLY|GLY|ME|A|RG|L|Y|GLY|A|RG|L|Y|GLY|L|E|U 260
DB      1718  CCGGAGTGTATGCTGCGCAAGAGGTGGCCCTGGTGGCAATGGCGGAGAAAGAGGTGGCTC 1777
QY      261  M|E|C|A|P|A|RG|L|Y|GLY|P|RO|GLY|GLY|ME|C|P|H|E|A|RG|L|Y|GLY|A|RG|L|Y|GLY|A|P|A|RG|L|Y|GLY 280
DB      1778  ATGATGCTGTGTGTCCTCCGTTGAATGTTCAAGAGTGGCCGTGTGAGACAGAGGTGC 1837
QY      281  P|H|E|A|RG|L|Y|GLY|A|RG|L|Y|ME|C|A|P|A|RG|L|Y|GLY|P|H|E|G|L|Y|GLY|A|RG|A|RG|L|Y|GLY|P|RO 300
DB      1838  TTCGTGTGTGGCCGGGCGCATGACCGAGGTGGCTTGTGTGAGAGAAACAGAGGTGCCT 1897
QY      301  G|Y|GLY|P|RO|P|RO|GLY|P|RO|L|E|U|ME|T|G|L|U|L|ME|G|L|Y|GLY|A|RG|A|RG|L|Y|GLY|A|RG|L|Y|GLY 320
DB      1898  GGGGGGGCCCTCTGACCTTTGATGAGAACATGAGAGAAAGAGAGACGATGAGGA 1957
QY      321  P|RO|GLY|Y|A|N|E|A|P|L|Y|S|E|G|I|U|H|A|RG|L|G|I|N|G|L|Y|A|RG|A|P|A|RG|P|RO|T|Y|R 338
DB      1958  CCTGGAATAATGATTAAGGAGACACGTCAGAGGCGAGATCGCCCTAC 2011

```

```

AC      ADS16277;
XX      18-NOV-2004 (first entry)
DT
XX
DE      Human CDNA encoding the Ewing sarcoma protein SegID 1.
XX
KW      human; sas; gene; Ewing sarcoma; EWS; prostatic cancer; alopecia; acne;
XX      hypogonadism; androgen-resistance syndrome; testicular feminisation.
XX
OS      Homo sapiens.
XX
FH      Key
FT      CDS
FT      44..2014
FT      /tag= a
FT      /product= "Ewing sarcoma protein"
XX
XX      EP145190-A1.
XX
XX      08-SEP-2004.
XX
XX      16-FEB-2004; 2004EP-00003422.
XX
XX      04-MAR-2003; 2003DB-01009280.
XX      25-APR-2003; 2003US-0465692P.
XX
XX      (SCHD ) SCHERING AG.
XX
XX      Obendorf M, Wolf S;
XX
XX      WPI: 2004-627861/61.
XX      DR
XX      P-PSDB; ADS16278.
XX
XX
XX      Determining the hormonal effects of substances, used to identify
XX      pharmaceuticals, e.g. for treatment of androgen receptor dysfunction,
XX      from modulating interaction between nuclear receptors and Ewing sarcoma
XX      protein.
XX
XX
XX      Claim 7, SEQ ID NO 1, 30pp; German.
XX
XX
XX      This invention relates to a novel modulators that alter the interaction
XX      between the Ewing sarcoma protein (EWS) and its nuclear receptor, as well
XX      as the screening method thereof. Specifically, it refers to determining
XX      and identifying a hormonal effect brought about by test compounds that
XX      modulate either the binding of EWS to the nuclear receptor or the ligand-
XX      induced activity of this receptor. The present invention describes the
XX      nuclear receptors as including oestrogen, progesterone, thyroid hormone,
XX      Vitamin D, and retinoic acid receptors, most preferably they are androgen
XX      receptors. Accordingly, these modulators may be used in the development
XX      of pharmaceutical compositions that can diagnose and be used to treat
XX      diseases associated with receptor dysfunction such as prostatic cancer,
XX      alopecia, acne, hypogonadism and androgen-resistance syndrome e.g.
XX      testicular feminisation. This method provides reliable, sensitive,
XX      simple, inexpensive and rapid assessment of the hormonal effects of these
XX      test compounds. This polynucleotide sequence is the CDNA encoding the
XX      human Ewing sarcoma protein of the invention.
XX
XX
XX      Sequence 2390 BP; 646 A; 589 C; 668 G; 487 T; 0 U; 0 Other;
XX
XX
XX      Alignment Scores:
XX      Pred. No.: 2, 14e-60 Length: 2390
XX      Score: 1922.00 Matches: 338
XX      Percent Similarity: 100.00% Conservative: 0
XX      Best Local Similarity: 100.00% Mismatches: 0
XX      Query Match: 100.00% Indels: 0
XX      DB: 13 Gaps: 0
XX
XX
XX      US-10-791-017A-2_COPY_319_656 (1-338) x ADS16277 (1-2390)
XX
QY      1  G|YGLYARGLYGLYMETGLYSER|AGLYGLYUARGLYGLY|PHEANLYSPROGLYGLY 20
DB      998  GGAGGAGCGCGGTGAATGGCGACCGCTGGAGAGGAGAGTGCTTCATTAAGCCCTGGTGA 1057
QY      21  P|MEC|A|SP|GLU|GLY|P|RO|A|P|LEU|A|P|LEU|GLY|P|RO|P|RO|VAL|A|P|RO|A|P|GLY|A|P|SER 40

```

```

Db      1058 CCATGATGATAGAGACCAAGTCTTGTATCTGAGCCCACTTAATCCAGTGAAGACTCT 1117
Qy      41  AspAsnSerAlaIleTyrValGlnGlyLeuAsnAspSerValThrLeuAspLeuAla 60
Db      1118 GACAAAGCTGCAATTATGTATCAAGAGATTAATGACAGTGTGACTGATGATGATCTGGCA 1177
Qy      61  AspPhePheLeuGlnCysGlyValValLeuMetAsnLysArgThrGlyGlnProMetIle 80
Db      1178 GACTTCTTTAAGAGAGTGTGGGTGTTAAGATGAACAAAGAACTGGGCAACCATGATC 1237
Qy      81  HisIleTyrLeuAspLysGlnThrGlyLysProLysGlyAspAlaThrValSerTyrGlu 100
Db      1238 CACATCTACCTGACAAAGAAACAGGAAACCCAAAGCCATGCCACAGTCTTATGAA 1297
Qy      101  AspProProThrAlaLysValAlaValGluTyrPheAspGlyLysAspPheGlnGlySer 120
Db      1298 GACCCACCCCACTGCGCAAGGCTGCGGTGAATGGTTGATGGGAAAGATTTCAGAGGAGC 1357
Qy      121  LysLeuLysValSerLeuAlaArgLysValProMetAsnSerMetArgGlyGlyLeu 140
Db      1358 AAACCTTAAAGTCTCCCTTGTCTGAGAGAGAGCTTCATTAAGATCGGGGTGCTG 1417
Qy      141  ProProArgGlnGlyArgGlyLysMetProProProLeuArgGlyGlyProGlyLys 160
Db      1418 CCACCCCGTGAAGGAGAGGAGCATGCCACCACTCCGTGAGAGTCCAGAGAGCCCAAGA 1477
Qy      161  GlyProGlyLysLysProMetGlyArgMetGlyGlyValArgGlyLysAspArgGlyLys 180
Db      1478 GGTCTGGGGAGACCATGGGTGCGATGGAGGCGGTGAGAGAGATAGAGAGGCTTCCTC 1537
Qy      181  ProArgGlyLysProArgGlySerArgGlyAsnProSerGlyLysGlyAsnValGlnHisArg 200
Db      1538 CCAGAGAGAGCCCGGGGTTCCTCCGAGGAGACCCCTCTGAGAGAGAAAGCTCCAGCACGA 1597
Qy      201  AlaGlyAspTyrGlnCysProAsnProGlyCysGlyValAsnGlnAsnPheAlaTyrArgThr 220
Db      1598 GCTGGAGACTGGCAGTGTCCCAATCCGGGTTGTGGAACCAAGAACTTCGCTGGAGAA 1657
Qy      221  GluCysAsnGlnCysValLysAlaProLysProGlnGlyPheLeuProProPheProPro 240
Db      1658 GAGTGCACACAGTGTAGAGCCCAAGCCCTGAAGGCTTCCTCCGCAACCCCTTCCGCC 1717
Qy      241  ProGlyLysAspArgGlyArgGlyLysLysProGlyLysMetArgGlyLysArgGlyLys 260
Db      1718 CCGGTGTGTATCTGTGCAAGAGTGTGCTGTGTGCAATGCGGGAGAGAAAGTGTGCTC 1777
Qy      261  MetAspArgGlyLysLysProGlyLysMetPheArgGlyLysArgGlyLysAspArgGlyLys 280
Db      1778 ATGATATGCTGTGTGTCCCGGTGGAATTTCAAGAGTGTGCGTGTGAGACAGAGCTGGC 1837
Qy      281  PheArgGlyLysLysArgGlyLysMetAspArgGlyLysPheGlyLysLysArgGlyLysPro 300
Db      1838 TTCCGTGTGTGCGCGGAGCATGAGCCGAGTGTGCTTGTGTGAGAGAAACAGAGTGTG 1897
Qy      301  GlyLysLysProProGlyLysProLeuMetGlnGlnMetGlyLysLysArgGlyLysArgGlyLys 320
Db      1898 GGGGGGGCCCTGTGACCTTTGATGAAACAGTGGAGAGAGAGAGAGAGAGCGTGGAGGA 1957
Qy      321  ProGlyLysMetAspLysGlyLysLysArgGlnGlnLysArgGlyLysAspArgGlyLys 338
Db      1958 CTGTGAGAAATGATTAAGCGGAGACACCGTCAAGAGCGGAGAGATCCGCCCTTAC 2011

```

RESULT 10

ACN40903 ID ACN40903 standard; cDNA; 2390 BP.

XX AC ACN40903;

XX DT 18-NOV-2004 (first entry)

XX DE Tumour-associated antigenic target (TAT) cDNA DNA269830, SEQ ID NO:5989.

KM Tumour-associated antigenic target; TAT; human; overexpression; cancer;
 KM tumour; diagnosis; cell proliferative disorder; breast cancer;
 KM colorectal cancer; lung cancer; ovarian cancer; liver cancer;
 KM central nervous system cancer; bladder cancer; pancreatic cancer;
 KM cervical cancer; melanoma; leukemia; hybridisation probe;
 KM chromosome identification; chromosome mapping; gene mapping;
 KM gene therapy; cytostatic; gene; ss.

OS Homo sapiens.
 XX MO2004030615-A2.
 XX 15-APR-2004.
 XX 29-SEP-2003; 2003WO-US028547.
 XX 02-OCT-2002; 2002US-0414971P.
 XX (GENTH) GENENTECH INC.
 XX Wu TD, Zhang Z, Zhou Y;
 XX WPI; 2004-347921/32.
 XX P-PSDB; ABM82330.
 XX New tumor-associated antigenic target polypeptides and nucleic acids,
 PT useful in preparing a medicament for treating or detecting a
 PT proliferative disorder, e.g. breast, lung, colorectal, ovarian or
 PT prostate cancer or tumor.
 XX Claim 1; SEQ ID NO 5989; 7273bp; English.

CC The invention relates to human tumour-associated antigenic target (TAT)
 CC polypeptides, and their related nucleic acids. The TAT polypeptides are
 CC overexpressed in cancer tissues compared to normal tissues, and may thus
 CC serve as effective targets for the diagnosis and treatment of cancer in
 CC mammals. The invention also relates to nucleic acid and polypeptide
 CC sequences at least 80% identical to the TAT nucleic acids and
 CC polypeptides; expression vectors and host cells comprising a TAT nucleic
 CC acid; an antibody specific for a TAT polypeptide; a peptide or organic
 CC molecule which binds to a TAT polypeptide; fusion proteins comprising a
 CC TAT polypeptide; and methods and compositions for the treatment or
 CC diagnosis of cancer in mammals. TAT polypeptides, nucleic acids,
 CC antibodies, antagonists, binding molecules and compositions are useful
 CC for diagnosing or treating a cell proliferative disorder associated with
 CC increased TAT expression, particularly cancers such as breast cancer,
 CC colorectal cancer, lung cancer, ovarian cancer, liver cancer, bladder
 CC cancer, pancreatic cancer, cervical cancer, cancers of the central
 CC nervous system, melanoma and leukemia. TAT nucleic acids may further be
 CC used as hybridisation probes, in chromosome and gene mapping, in
 CC chromosome identification and in gene therapy. The present sequence
 CC represents a TAT nucleic acid of the invention.

SQ Sequence 2390 BP; 645 A; 589 C; 668 G; 488 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	Length:	Matches:	Mismatches:	Indels:	Gaps:
2, 14e-60	2390	338	0	0	0
Score:	1922.00	Conservative:	0		
Percent Similarity:	100.00%				
Best Local Similarity:	100.00%				
Query Match:	100.00%				

US-10-791-017A-2_COPY_319_656 (1-338) x ACN40903 (1-2390)

```

Qy      1  GlyLysArgGlyLysLysMetGlySerAlaGlyLysArgGlyLysPheAsnLysProGlyLys 20
Db      998 GAGAGAGCGGTGTGAATGGGCAAGCGCTGGAGAGCGAGTGTGCTTCAATAAGCCGTGGGA 1057
Qy      21  ProMetAspGlnGlyLysProAspLeuAspLeuGlyLysProValAspProAspGlyAspSer 40
Db      1058 CCATGATGATAGAGACCAAGTCTTGTATCTGAGCCCTCTGTATGATCCAGTGAAGACTCT 1117

```

QY 41 AspAsnSerAlaIleTyrValGlnGlyLeuAsnAspSerValThrLeuAspLeuAla 60
 DB 1118 GACAAAGAGTCATTTATGTAACAAGATTAAATGACGTGTGACTTAAAGTATGATCTGGCA 1177
 QY 61 AspPhePheIleGlnGlyValValIleMetAsnIleArgThrGlyGlnPrometIle 80
 DB 1178 GACTTCTTTAAGACGTGTGGGTTGTTAAGATGAACAAGAACTGGGCAACCATGATC 1237
 QY 81 HisIleTyrLeuAspIleGlyLeuThrGlyValAspProIleValAspIleThrValIleTyrGlu 100
 DB 1238 CACATCTACCTGAGCAAGAAACAGGAAACCCAAAGGCGATGCCACAGTGTCTATGAA 1297
 QY 101 AspProProThrAlaIleValAlaValGluThrPheAspGlyValAspPheGlnGlySer 120
 DB 1298 GACCCACCCCACTCCCAAGGCTCCGTGAATGGTTGATGGAAAGATTTTCAAGGGAGC 1357
 QY 121 LysLeuIleValSerLeuAlaArgIleValPheProPheMetAsnSerMetArgIleValLeu 140
 DB 1358 AAACCTTAACTCTCCCTGCTCCGAAAGACCTCCATATGAACTATGCGGGTGTCTG 1417
 QY 141 ProProArgGluIleValArgIleMetProProProLeuArgIleGlyProGlyGlyProGly 160
 DB 1418 CCAACCCCGTGAAGGAGGATGCAACCACTCCGTGAGGTCCAGAGGCGCCAGGA 1477
 QY 161 GlyProGlyValProMetGlyValArgIleValArgIleValAspArgGlyValPhePro 180
 DB 1478 GGTCTGGGGGAGCCATGGGTCCATGGGAGGCGGTGAGAAATGAGAGGCTTCCCT 1537
 QY 181 ProArgGlyProArgGlySerArgGlyValAsnProSerGlyValGlyValAsnValGlnHisArg 200
 DB 1538 CCAGAGAGACCCCGGGGTTCCCAAGGAAACCTCTGGAAGAGAAAGCTTCACAGACCA 1597
 QY 201 AlaGlyAspIleGlnGlyValProAsnProGlyValGlyValAsnGlnAsnPheAlaIleArgThr 220
 DB 1598 GCTGAGAGACTGGCAGGTGCTCCCAATCCGGGTTGGAAACCAAGAACTTCCGTGAGAGCA 1657
 QY 221 GluCysAsnGlnCysValAlaIleProIlePheLeuProProPheProPro 240
 DB 1658 GAGTGCACCAAGTGAAGGCCCAAGGCTGAAAGGCTTCCCGCCACCTTCCGCC 1717
 QY 241 ProGlyValAspArgGlyValArgIleValPheArgIleValArgIleValLeu 260
 DB 1718 CCGGGTGTGATCTGTGCAAGGTGGCCCTGTGTGGCATGGGGAGAGAAAGGTGGCTTC 1777
 QY 261 MetAspArgGlyValIleProGlyValMetPheArgIleValArgIleValAspArgGlyVal 280
 DB 1778 ATGATCTGTGTGTGCTCCGGTGAATGTTCAAGAGTGGCGGTGTGAGACAGAGGTGGC 1837
 QY 281 PheArgGlyValArgIleValMetAspArgGlyValPheGlyValValArgArgIleValPro 300
 DB 1838 TTCCGTGTGTGGCCGGGCGATGAGACGAGGTGGCTTGTGTGAGAGAAAGCAAGGTGGCCCT 1897
 QY 301 GlyValProProGlyValProLeuMetGlnGlnMetGlyValValArgArgIleValArgGlyVal 320
 DB 1898 GGGGGGGCCCCCTGATCTTGTATGTAACAATGGAGAAAGAAAGAGAGGACGTTGAGGA 1957
 QY 321 ProGlyValMetAspIleGlyValHisArgGlnGlnArgArgAspArgProTyr 338
 DB 1958 CCTGGAATAATGATTAAGCGAGCACCTTCAGAGAGGCGAGAGTGGGCCCTTAC 2011
 RESULT 11
 ADRO7446
 ID ADRO7446 standard; cDNA; 2026 BP.
 XX
 AC ADRO7446;
 XX
 DT 04-NOV-2004 (first entry)
 XX
 DE Full length human cDNA useful for treating neurological disease Seq 952.
 XX
 KW Gene; 86; human; oligo-capping method; diagnostic marker; gene therapy;
 KW osteoporosis; neurological disease; Alzheimer's disease;
 KW Parkinson's disease; dementia; short memory; cancer;

KW sense or motor function; emotional reaction; fear response; panic;
 KW osteopathic; neuroprotective; nootropic; antiparkinsonian; cyostatic;
 KW tranquilliser.
 XX
 OS Homo sapiens.
 XX
 PN EP1447413-A2.
 XX
 PD 18-AUG-2004.
 XX
 PF 12-FEB-2004; 2004EP-00003145.
 XX
 PR 14-FEB-2003; 2003JP-00102207.
 PR 09-MAY-2003; 2003JP-00131452.
 XX
 PA (REAS-) RES ASSOC BIOTECHNOLOGY.
 XX
 PI Isogai T, Yamamoto J, Nishikawa T, Isono Y, Sugiyama T, Otsuki T,
 PI Wakamatsu A, Ishii S, Nagai K, Irie R;
 DR WPI; 2004-583265/57.
 DR P-PSDB; ADRO9402.
 XX
 PT New 1995 cDNA, useful for treating osteoporosis, neurological diseases,
 PT Alzheimer's diseases, Parkinson's diseases, dementia and various cancers.
 XX
 PS Claim 1; SEQ ID NO 952; 2686pp; English.
 XX
 CC This invention relates to novel, isolated full length human cDNA
 CC molecules and the encoded proteins thereof. Specifically, it refers to
 CC cDNA clones obtained by an oligo-capping method, where none of these
 CC clones are identical to any known human mRNAs. The present invention
 CC describes an immunoassay to identify agonists and antagonists, as well as
 CC antibodies, antisense molecules and siRNAs that can all be used to bind
 CC to and modulate expression of the cDNA molecules. As such, these
 CC molecules are useful for diagnostic markers or therapeutic targets for
 CC the various diseases or morbid states. In particular, they are useful in
 CC gene therapy for treating osteoporosis, neurological disease, Alzheimer's
 CC disease, Parkinson's disease, dementia, short memory and various cancers,
 CC as well as for maintaining equilibrium of sense or motor function, and
 CC for treating emotional reaction, fear response and panic. Accordingly,
 CC they exhibit osteopathic, neuroprotective, nootropic, antiparkinsonian,
 CC cyostatic and tranquilliser activities. This polynucleotide is a full
 CC length human cDNA sequence of the invention. NOTE: This sequence is not
 CC given in the sequence listing of the specification but can be obtained on
 CC CD-ROM from the European Patent Office, Vienna Sub-office.
 XX
 SQ Sequence 2026 BP; 540 A; 513 C; 590 G; 383 T; 0 U; 0 Other;
 Alignment Scores:
 Pred. No.: 5,53e-60 Length: 2026
 Score: 1909.00 Matches: 337
 Percent Similarity: 99.70% Conservative: 0
 Best Local Similarity: 99.70% Mismatches: 1
 Query Match: 99.32% Indels: 0
 DB: 13 Gaps: 0
 US-10-791-017A-2_COPY_319_656 (1-338) x ADRO7446 (1-2026)
 QY 1 GlyValArgGlyValMetGlySerAlaGlyValArgGlyValPheAsnIleProGlyVal 20
 DB 836 GAGAGACGGGTGAATGGCAGCGCTGAGAGCGAGGTGGCTTCAATAAAGCCTGTGGGA 895
 QY 21 ProMetAspGluIleValProAspLeuAspLeuGlyProProValAspProAspGluAspSer 40
 DB 896 CCAATGATGAAGAGCAAGATCTTGAATCTAAGGCCACCTGTATATCCAAGTAAAGACTCT 955
 QY 41 AspAsnSerAlaIleTyrValGlnGlyLeuAsnAspSerValThrLeuAspLeuAla 60
 DB 956 GACAAAGAGTCATTTATGTAACAAGATTAAATGACAGTGTGACTTAAAGTATGATCTGGCA 1015
 QY 61 AspPhePheIleGlnGlyValValIleMetAsnIleArgThrGlyGlnPrometIle 80

Db 1016 GACTTCTTAAAGCAGTGTGGGTTGTATTAGATGAAACAGAACTGGGCAACCATGATC 1075
 Qy 81 Hs11eTyLeuAspIyRgIyUthRgIyLysProIyGlyYAspAlaThrValSerTyRgIu 106
 Db 1076 CACATCTACCTGAGCAAGAAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1135
 Qy 101 AspProProthralaIyValaIaValaIaGluTrpPheAspIyLysAspPheGlnGlySer 120
 Db 1136 GACCCACCCACCTCCAG 1195
 Qy 121 LysLeuLysValSerLeuAlaArgIyLysProPheAspSerMetArgIyGlyLeu 140
 Db 1196 AAACCTTAAGTCTCCCTTGTCTCGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1255
 Qy 141 ProProArgIyGluIyArgIyMetProProProLeuArgIyGlyProGlyIyProGly 160
 Db 1256 CCAACCCGAG 1315
 Qy 161 GlyProGlyYAspIyProMetGlyArgMetGlyYArgGlyYAspArgGlyIyPhePro 180
 Db 1316 GGTCTCTGGGAG 1375
 Qy 181 ProArgIyProArgIySerArgIyAspProSerGlyYArgIyLysValaGlnHsArg 200
 Db 1376 CCAAG 1435
 Qy 201 AlaGlyYAspTrpGlnIyAspProAspProGlyYArgGlyYAspGlnHsPheAlaTrpArg 220
 Db 1436 GCTGAG 1495
 Qy 221 GlyCyAspGlnIyCysLysValaProIyPheProGlyYArgGlyYAspPheProPro 240
 Db 1496 GAGTGCACACAGTGTAG 1555
 Qy 241 ProGlyYAspArgGlyArgGlyYArgIyProGlyYMetArgGlyYArgGlyYLeu 260
 Db 1556 CCGAGTGTGTATCGTGGCAG 1615
 Qy 261 MetAspArgGlyYArgIyProGlyYArgIyMetPheArgGlyYArgIyYAspArgGlyY 280
 Db 1616 ATGATCTGT 1675
 Qy 281 PheArgGlyYArgIyMetAspArgGlyYArgIyPheGlyYArgIyYAspArgGlyYPro 300
 Db 1676 TTCCTGT 1735
 Qy 301 GlyYLeuProProGlyYProLeuMetGlnGlnMetGlyYArgIyArgGlyYArgGlyY 320
 Db 1736 GGG 1795
 Qy 321 ProGlyYMetMetAspIyGlyYArgIyGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 338
 Db 1796 CCTGAGAAATGATTAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1849
 RESULT 12
 ABI99383
 ID ABI99383 standard; cDNA; 2188 BP.
 XX
 AC ABI99383;
 XX
 DT 07-MAR-2002 (first entry)
 XX
 DE Mouse ischaemic condition related cDNA sequence SEQ ID NO:288.
 XX
 KW Mouse; ischaemia; compressive ischaemia; occlusive ischaemia;
 KM vasospastic ischaemia; ischaemic condition; ischaemic disease; ss.
 XX
 OS Mus musculus.
 XX
 FN M0200188188-A2.
 XX
 PD 22-NOV-2001.
 XX

PF 18-MAY-2001; 2001WO-0P004192.
 XX
 PR 18-MAY-2000; 2000JP-00145977.
 XX
 PA (UNNT-) UNIV NIHON SCHOOL JURIDICAL PERSON.
 XX
 PI Ishikawa K, Arai S, Takahashi Y, Nagata T, Ishii Y;
 XX
 DR MPI; 2002-034733/04.
 DR P-PSDB; ABB57126.
 XX
 PT Examining the ischemic condition (e.g. occlusive ischemia) by measuring
 PT expression levels of particular genes defined in the specification or by
 PT determining the expression profile of a gene group comprising these
 genes.
 XX
 PS Claim 2; Page 794-799; 2690pp; English.
 XX
 CC The present invention describes a method for examining ischaemic
 CC conditions, comprising measuring the expression levels of particular
 CC genes (1) in a test sample or determining the expression profile of a
 CC gene group in the sample comprising genes selected from (1). The method
 CC is useful for examining the ischaemic condition (e.g. compressive
 CC ischaemia, occlusive ischaemia or vasospastic ischaemia) by measuring
 CC expression levels of particular genes (ABI99202 to ABI99912, encoding the
 CC protein sequences in ABB57020 to ABB57374) or by determining the
 CC expression profile of a gene group comprising these genes. The expression
 CC levels or expression profiles produced by these genes are used as an
 CC indicator when screening for ischaemic condition-improving drugs or
 CC therapeutics for ischaemic diseases. ABI99913 and ABI99914 represent PCR
 CC primers for a mouse ischaemic condition related sequence, which are used
 CC in the exemplification of the present invention
 XX
 SQ Sequence 2188 BP; 587 A; 551 C; 638 G; 412 T; 0 U; 0 Other;
 Alignment Scores:
 Pred. No.: 3,12e-59 Length: 2188
 Score: 1888.50 Matches: 332
 Percent Similarity: 99.41% Conservative: 4
 Best Local Similarity: 98.22% Mismatches: 1
 Query Match: 98.26% Indels: 1
 DB: 6 Gaps: 1
 US-10-791-017a-2_copy_319_656 (1-338) x ABI99383 (1-2188)
 Qy 1 GlyYArgGlyYArgIyMetGlySerAlaGlyYArgGlyYArgIyPheAspIyProGlyY 20
 Db 1011 GAG 1067
 Qy 21 PrometAspGlnGlyProAspLeuAspLeuGlyProProValaAspProAspGluAspSer 40
 Db 1068 CCAATGATGAG 1127
 Qy 41 AspAspSerAlaIleTyValaGlnGlyLeuAspSerValThrLeuAspAspLeuAla 60
 Db 1128 GACAAAGAGCAATTTATGTGACAGAGATTAAATGACAAATGTGATGTATGTGGCA 1187
 Qy 61 AspPhePheLysGlnCysGlyValaValaLysMetAspLysArgThrGlyGlnProMetIle 80
 Db 1188 GACTTCTTAAAGCAGTGTGGGTTGTCAAGATGACAAAGAGAGAGAGAGAGAGAGAGAG 1247
 Qy 81 Hs11eTyLeuAspIyRgIyUthRgIyLysProIyGlyYAspAlaThrValSerTyRgIu 100
 Db 1248 CACATCTACCTGAGTAAG 1307
 Qy 101 AspProProthralaIyValaIaValaIaGluTrpPheAspIyLysAspPheGlnGlySer 120
 Db 1308 GATCCACCAACTGCAG 1367
 Qy 121 LysLeuLysValSerLeuAlaArgIyLysProPheAspSerMetArgIyGlyLeu 140
 Db 1368 AAACCTTAAGTCTCCCTTGTCTCGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1427

QY	141	PROF	PROA	RGSLUG	LA	ARGGL	WME	PRO	PRO	PRO	LEU	AR	VG	GL	PRO	GL	160
Db	1428	CCACCT	CGTGA	GGG	CGA	GGGG	GTAT	GC	CA	CC	CA	CA	CTT	CG	TGA	GG	1487
QY	161	GL	PRO	GL	Y	PRO	ME	GL	Y	ARG	ME	GL	Y	GL	Y	ARG	180
Db	1488	GGCCCT	TGG	AG	GA	GC	CA	TGG	GT	GC	AT	TGG	GA	GG	CC	CTT	1547
QY	181	PRO	ARG	GL	Y	PRO	ARG	GL	Y	SE	ARG	GL	Y	ASN	P	OS	200
Db	1548	CCAA	AG	GG	GG	CCCC	CG	AG	GG	CT	CC	CA	GA	AA	CCCC	CT	1607
QY	201	ALA	GL	Y	ASP	TR	PG	LI	NC	Y	PRO	ASN	PRO	GL	Y	CY	220
Db	1608	GCT	G	AG	AC	TG	CA	GT	GT	CC	CA	AT	CC	GG	AG	CT	1667
QY	221	GL	Y	CA	BA	AN	GL	NC	Y	ALA	AP	RO	LY	SP	PRO	GL	240
Db	1668	GAA	T	CA	CA	AC	CA	CT	TA	GA	GG	CCCC	CT	TA	AG	CC	1722
QY	241	PRO	GL	Y	ASP	ARG	GL	Y	ARG	GL	Y	PRO	GL	Y	ME	ARG	260
Db	1728	CCG	GT	GT	GT	GT	AT	CG	TG	CA	AG	GT	GG	CC	CT	GT	1787
QY	261	ME	TA	PA	RG	GL	Y	GL	Y	PRO	GL	Y	ME	TA	RG	GL	280
Db	1788	AT	GA	CC	CG	GT	GT	GT	CT	CG	TA	GA	AA	TG	TT	CA	1847
QY	281	PH	EA	RG	GL	Y	ARG	GL	Y	ME	TA	RG	GL	Y	PH	EA	300
Db	1848	TT	CC	GA	GG	TT	GG	CC	CG	TA	TA	GA	CC	GA	GG	TT	1907
QY	301	GL	Y	GL	Y	PRO	GL	Y	PRO	LEU	ME	GL	GL	Y	ME	GL	320
Db	1908	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	1966
QY	321	PRO	GL	Y	ME	TA	RG	GL	Y	GL	Y	HA	RG	GL	Y	HA	338
Db	1968	C	T	G	G	G	A	A	A	A	A	A	A	A	A	A	2021

RESULT 13

ADP56333

ADP56333 standard; cDNA, 1807 BP.

ADP56333;

18-NOV-2004 (first entry)

Human PRO cDNA sequence SEQ ID NO:2309.

human; PRO; immune related disease; inflammatory immune response; immune response stimulation; antiallergic; antianemic; antilathritic; antihaematic; antidiabetic; antiinflammatory; antipneumatic; antirheumatic; anticholelith; CNS; dermatological; gastrointestinal; haemostatic; hepatotropic; immunostimulant; immunosuppressive; muscular; nephrotropic; neutropotective; osteopathic; respiratory; vasotropic; virocidic; gene therapy; gene; ss.

Homo sapiens.

WO2004039956-A2.

13-MAY-2004.

28-OCT-2003; 2003WO-US034381.

29-OCT-2002; 2002US-0422472P.

(GETH) GENENTECH INC.

Aggarwal S, Clark H, Gurney AL, Schoenfeld J, Williams PW; Wood WT, Wu TD;

DR WPI: 2004-376182/35.
DR P-PSDB; ADP56334.
XX
XX New PRO polynucleotides and polypeptides, useful in diagnosing
PT and treating an immune related disease, e.g. systemic lupus
PT erythematous, rheumatoid arthritis, diabetes mellitus or asthma and in
PT stimulating an immune response.
XX
XX
XX Claim 2; SEQ ID NO 2309; 3009PD; English.
XX
XX
XX The present invention describes an isolated PRO nucleic acid (1). Also
CC described: (1) a vector comprising (1); (2) a host cell comprising the
CC vector of (1); (3) a process for producing a PRO polypeptides; (4) an
CC isolated PRO polypeptide; (5) a chimeric molecule comprising the
CC polypeptide of (4) fused to a heterologous amino acid sequence; (6) an
CC antibody which specifically binds to a polypeptide of (4); (7) a
CC composition of matter comprising a polypeptide of (4), an agonist or
CC antagonist of the polypeptide or an antibody that binds to the
CC polypeptide in combination with a carrier; (8) an article of manufacture
CC comprising a container, a label on the container and a composition of
CC matter of (7); (9) a method of treating an immune related disease in a
CC mammal; (10) a method for determining the presence of a PRO polypeptide
CC in a sample suspected of having the polypeptide; (11) a method of
CC diagnosing an immune related disease or an inflammatory immune response
CC in mammal; (12) a method of identifying a compound that inhibits or
CC mimics the activity of or expression of a gene encoding a PRO polypeptide
CC and (13) a method of stimulating the immune response in a mammal. The
CC PRO sequences have anti-allergic, anti-nausea, anti-arthritis,
CC antiasthmatic, antidiabetic, anti-inflammatory, antiparathetic,
CC antirheumatic, antihypoid, CNS, dermatological, gastrointestinal,
CC haemostatic, hepatotropic, immunostimulant, immunosuppressive, muscular,
CC nephrotropic, neuroprotective, osteoplastic, respiratory, vasotropic and
CC virucide activities, and can be used in gene therapy. The nucleic acid
CC (1) and the encoded polypeptides, compositions, kits and methods are
CC useful in diagnosing and treating an immune related disease and in
CC stimulating an immune response. The present sequence represents a human
CC PRO nucleotide sequence from the present invention.
XX
XX
XX Sequence 1807 BP; 476 A; 497 C; 523 G; 311 T; 0 U; 0 Other;
SQ
XX
XX
XX Alignment Scores:
Pred. No.: 1.68e-56 Length: 1807
Score: 1810.00 Matches: 318
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 90.00% Mismatches: 0
Query Match: 14.17% Indels: 0
DB: 13 Gaps: 0
US-10-791-017A-2_COPY_319_656 (1-318) X ADP56333 (1-1807)
QY 21 PrometxapnglugslyproaspLeuaspLenglyProproValaaspProaspLyspSer 40
DB 851 CCAGTGGATGAGGACCAAGATCTTGAAGTGGAGGCTCTGTAAGATCCAGATGAAGACTCT 910
QY 41 AspaanserajalietyrValgInglyLeuaspAspserValThrLeuaspAspLeuA 60
DB 911 GACAAACAGTGCATTTATATGTAACAAGATTAATGACAGTGTGACTTAAGATGATCTGCA 970
QY 61 AsphepelysgIncygLyValVallybMetasnllysarGthrGlyInPrometile 80
DB 971 GACTTCCTTAAAGCAGTGTGGGGGTGTTAAGATGTAACAAGAACCTGGCAACCATGATC 1030
QY 81 HistletyrieuaspLysgluThrGlylybProcygLyaspAlatrpValserTyrglu 100
DB 1031 CACATCTACCTGGCAAGAAACGAAACGAAAGCCCAAGAGCCAGTGCACAGTCTTATGAA 1090
QY 101 AspProcthrAlatlyalaalaValgIntrpPheaspGlylyAspPhegInglySer 120
DB 1091 GACCAACCACTGCACAAAGCGCTCGCTGGAAATGCTTTATGGGAAAGATTTTCAAGGAGC 1150
QY 121 LysleuLysValserLeuAlaargLylybProProMetasnsrMetarGlylyLeu 140
DB 1151 AAACCTTAAAGCTCCCTGGCTCGGAAAGAGCTCTCAATGAAACAGTATGCGGGGTGCTG 1210

Qy	141	troProAaTgGluGlyAaTgGlyWaeProProProleuAaTgGlyVgLyProGlyVgLyProGly	160
Db	1211	CCACCCCGTGAAGGCGAGAGGCAATGCCACCACTCCGTGAGGTCCAGAGGCCCAAGA	12707
Qy	161	GlyProGlyVgLyProMetGlyAaTgMetGlyVgLyAaTgGlyVgLyAaPaTgGlyVgLyPhePro	180
Db	1271	GGTCTCGGAGGAGCCCATGGGTTCGATGGAGAGGCCGTGAGAGGATPAGAGAGCTTCCCT	13308
Qy	181	ProAaTgGlyProAaTgGlyVseTArGlyVasnPProSerGlyVgLyVgLyAasVAlGlnHiaTg	200
Db	1331	CCAAGAGGACCCCGGGGTTTCCCGAGAGAAACCTCTTGAAGAGAGAAAGCTCAGACCGA	1390
Qy	201	AlaGlyAaPTTPGlnCyAProAaNPProGlyCySGlyAaNGlnAaNPheAlATrPaTgThr	220
Db	1331	GCTGGAGACTGGCACTGTCTCCAACTCCGGGTGTGTGAACCAAGAACTTGCTCGAGAGCA	1456
Qy	221	GlUCyAaNGlnCyAelYalAaProLybProGluGlyPheLeuProProProPhePro	240
Db	1451	GAGTGAACAACAGTGTAAAGGCCCAAGCCTGAAGGCTTCTCCGCAACCTTTCGGCC	1510
Qy	241	ProGlyVgLyAaPaTgGlyVgAaTgGlyVgLyProGlyVgLyMetAaTgGlyVgLyAaGlyVgLyPhe	260
Db	1511	CCGGGTGTGTATCGTGCAAGAGTGGCCCTGTGTGCATGCGGGAGAGAAAGTGGCTC	15707
Qy	261	MetAaPaTgGlyVgLyProGlyVgLyWaePheAaTgGlyVgLyAaTgGlyVgLyAaPaTgGlyVgLy	280
Db	1571	ATGATCTGTGTGTGTCTCCCGTGTGAATGTTCAAGAGTGGCCGTGTGTGAAGACAGAGTGGC	16308
Qy	281	PheAaTgGlyVgLyAaTgGlyWaeAaPaTgGlyVgLyPheGlyVgLyVgLyAaTgGlyVgLyPro	300
Db	1631	TTCCCTGTGTGGCCGGGCAATGACCGAGGTGTGTTGTGTGAAGAGCAAGTGGCCCT	1690
Qy	301	GlyGlyProProProGlyProLeuWaeGluGlnMetGlyVgLyAaTgAaTgGlyVgLy	320
Db	1631	GGGGGGGCCCCCTGGACTTTGATGGAAAGATGGAGAGAGAGAGAGACGTGGAGGA	1750
Qy	321	ProGlyLyWaeMetAaPlyAaGlyVgLyWHiAaTgGlnGluAaTgAaPaTgProTyr	338
Db	1751	CCTGGAAAATAAGATAAAGAGGAGCACCGTCAAGAGCCAGAGATGGCCCTTAC	1804
RESULT 14			
ID	AAS62623/c		
XX	AAS62623 standard; cDNA; 2273 BP.		
XX	AAS62623;		
DT	14-FEB-2002 (first entry)		
DE	cDNA sequence #410 encoding novel human secreted protein.		
KW	Human secreted protein; hyperproliferative disorder; autoimmune disorder;		
KW	immune deficiency disorder; blood disorder; inflammatory disorder;		
KW	infectious disorder; gene therapy; antimicrobial; hepatotropic;		
KW	immunosuppressive; antirheumatic; ss.		
OS	Homo sapiens.		
PN	WO200177291-A2.		
PD	18-OCT-2001.		
PF	29-MAR-2001; 2001WO-US010485.		
PR	06-APR-2000; 2000US-0195604P.		
PA	(GENY) GENETICS INST INC.		
PI	Wong GG, Clark HF, Fechtel K, Agostino MJ, Howes SH, Resnick RJ,		
DR	Gulikova K, Graham JR;		
XX	WPI; 2002-010900/01.		

PT	New polynucleotides encoding secreted proteins useful for treating e.g.
PR	asthma, HIV and Crohn's disease.
PS	Claim 1; Page 291; 391pp; English.
XX	
CC	The present invention relates to the isolation of novel cDNA sequences
CC	which encode human secreted proteins. The cDNA sequences have been
CC	derived from a variety of human tissues. The invention also provides a
CC	method for producing proteins from these polynucleotide sequences. The
CC	proteins are useful for identifying compounds that modulate their
CC	activity and production, and the cell is also useful for identifying
CC	compounds that modulate expression of the polynucleotide sequences
CC	encoding the secreted proteins. The sequences of the invention are useful
CC	for treating diseases such as hyperproliferative disorders (e.g., cancer),
CC	immune deficiency disorders (e.g., severe combined immunodeficiency
CC	(SCID)), autoimmune disorders (e.g., multiple sclerosis), blood disorders
CC	(e.g., thrombocytopaenia), inflammatory disorders (e.g., arthritis) and
CC	infectious disorders (e.g., hepatitis). The polynucleotide sequences of
CC	the invention are also useful in gene therapy. AA62214-AA62838
CC	represent the cDNA sequences of the invention that encode for novel human
CC	secreted proteins
XX	
SQ	Sequence 2273 BP; 477 A; 636 C; 562 G; 598 T; 0 U; 0 Other;
Alignment Scores:	
Pred. No.:	2,93e-55 Length: 2273
Score:	1777.00 Matches: 319
Percent Similarity:	94.38% Conservative: 0
Best Local Similarity:	94.38% Mismatches: 1
Query Match:	92.46% Indels: 18
DB:	Gaps: 2
US-10-791-017A-2_COPY_319_656 (1-338) x AA62623 (1-2273)	
OY	1 G1yG1yATgG1yG1yMeTg1ySeRAlG1yGUARg1yG1yPheASnLyPrOG1yG1y 20
Dd	1312 GGAAGAGCGCGGTGGAAATGGGC---GCTGGAAGAGAAGGTGGCTTCATATAGCTGTGCA 1256
OY	21 PwMcEApBg1UG1yPrOApLeuAspLeuGlYPrOPvOlAapPrOAsPg1UAspSer 40
Dd	1255 CCCATGTGTGAAGACCAAGACTTGATCTTAGGCCACCCTGTATGATCCAGATGAMAACCT 1196
OY	41 AspAmSerAlAlEtYrVAlNG1yLeuAsnAspSerValThrLeuAspLeuLA 60
Dd	1195 GACAACAGTGAATTATATGACAAAGATTAATGACAGTGGACTTATGATGATTTGGCA 1133
OY	61 AspPhePhelYSG1NCySG1yValVallyMecASnLYSAFTNhg1yG1NPrometIle 80
Dd	1135 GACTTCTTTAAGCAGTGTGGGGCTTGTATAGATGAAACAAGAAACCTGGGCAACCATGATC 1076
OY	81 HisIleTyrlEuAspLyG1UrTh-G1yLyePrOlySG1yAspAlArTh-VAlSerTYrgLU 100
Dd	1075 CACATCTAACCTGGCAAGAAACAGAAAGCCCAAAGCGATGCCACAGTGTCTTATGAA 1016
OY	101 AspProPOTMrAlALySAlAlAVAlGLUrTPheAsPg1yLYAspPhGlnGlySer 120
Dd	1015 GAACCAACCACTGCGCAAGGCTGCCGTGGAAATGTTGTGGAAAGATTTTCAGAGAGC 956
OY	121 LysLeuLySAlSerLeuAlARGLyLysPPrOmEcAnSermETArg1yG1yLeu 140
Dd	955 AAACTTAAAGTCTCCCTTGTCTCGAAGAAAGCTCCAAATGAACAGTATCGGGGTGTCTG 896
OY	141 ProBROArG1UG1yARG1yMecPrOBROPrOLEUAArg1yG1yPrOG1yG1yPrOG1y 160
Dd	895 COACCCGCTGAGGGGAGAGGATGCCAACCACTCCGTGAGGCTCCAGAGAGCCCAAGGA 836
OY	161 GLyPrOG1yG1yPrOMeTG1yARgmEcGLyLARg1yG1yAsPARg1yG1yThePro 180
Dd	835 GGTCCTGGGGAGACCATGGTGTGCAATGGAGGCGGTGAGAGATAGAGAGGCTTCCCT 776
OY	181 ProARg1yPrOAArg1ySERArg1yASnPROserG1yG1yASnVALGlnHisArg 200
Dd	775 CGAAGAGGAGCCCGGGGTTCCCGAGAGGAAACCTCTTGAGAGGAAAGCTCCAGCACCGA 716

FT /*tag= ag
 FT /number= 17
 FT polyA_signal 3102..33107
 FT /tag= an
 FT polyA_signal 3289..33294
 FT /tag= at
 XX
 PN W09323549-A2.
 XX
 PD 25-NOV-1993.
 XX
 PF 19-MAY-1993; 93MO-FR000494.
 XX
 PR 20-MAY-1992; 92FR-00006123.
 XX
 PA (CNRS) CNRS CENT NAT RECH SCI.
 XX
 PI Aurias A, Delattre O, Desmaze C, Meiot T, Peter M, Plougastel B,
 PI Thomas G, Zucman J,
 XX
 DR WPI; 1993-386580/48.
 DR P-PSDB; AAR44555.
 XX
 PT New nucleic acid of EWS gene and its hybrid(s) - contg. gene sequence
 PT involved in chromosomal translocation, also derived mRNA, probes, fusion
 PT proteins etc., for diagnosis and treatment of Ewing sarcoma and melanoma.
 XX
 PS Disclosure; Fig 6 and Fig 12; 123pp; French.
 XX
 CC The intron-exon junctions of the human Ews gene and the Hum-F11-1 gene
 CC have been sequenced (see AAQ50646 and AAQ50662, respectively). The
 CC different fusion products could be formed by fusing exons from the
 CC two genes, as happens after specific chromosomal translocations, can be
 CC predicted (see AAQ50671-Q50678). The sequences at fusion junctions of
 CC other observed translocations are given in AAQ50679-Q50683). NOTE: the
 CC 3309 residue sequence AAQ50646 does not appear in the specification; it
 CC is a combination of the Ews cDNA sequence (Fig 6, AAQ50643) and the
 CC individual intron sequences with their intron-exon junctions (Fig 12).
 CC (Updated on 25-MAR-2003 to correct PN field.)
 XX
 SQ Sequence 3309 BP; 822 A; 775 C; 857 G; 855 T; 0 U; 0 Other;
 SQ
 Alignment Scores:
 Pred. No.: 6.68e-54 Length: 3309
 Score: 1742.00 Matches: 338
 Percent Similarity: 65.25% Conservative: 0
 Best Local Similarity: 65.25% Mismatches: 0
 Query Match: 90.63% Indels: 180
 Gaps: 9
 DB:
 US-10-791-017a-2_COPY_319_656 (1-338) x AAQ50646 (1-3309)
 QY 1 G1YGLYARGLYGLYMETGLY----- 7
 DB 1398 GGAAGAGCGGTGGTAATGGGTAAAGCAACTTTCTCTTACCTTAAGGCTTCA 1457
 QY 8 -----SerAlaG1YGLYARGLYGLYPheanLysProGLY--- 19
 DB 1458 TTTCTCGTTATCCCGCCAGCAGCGCTGGAGAGCGAGGTGCTTCAATAAGCCGTGGT 1517
 QY 20 -----G1Y 20
 DB 1518 AAGTTTGTGATTATCATAGATAGATATTTTATATGATCTTCTCGTGGCGAGGA 1577
 QY 21 ProMetAspGluGlyProAspLeuAspLeu----- 30
 DB 1578 CCCATGATGATGAAGACCAAGATCTGATCTAGTAAGTGAATTCCTAGTTGGCTTCA 1637
 QY 31 -----GlyProProValAspProAspGluAspSer 40
 DB 1638 TATATATCTCTCTGTTGTTGTCTCTGAAAGGCCCACTGATGATCCAGATGAAGACTCT 1697
 QY 41 AspAsnSerAlaIleTyrValGlnGlyLeuAsnAspSerValThrLeuAspAspLeuAla 60

DB 1698 GACAAAGTGCATTTATATATACAGATTAATAATGACAGTGTGCTCTAGATGATCTGGCA 1757
 QY 61 AspPhePheLysGlnCysGlyValValLys----- 70
 DB 1758 GACTTCTTAAACAGTGGGGTGTGTTAAGGTCACTAAAGCATTAACAGGTCATCTGGC 1817
 QY 71 -----MetAsnLysArgThrGlyGlnProMetIle 80
 DB 1818 TCATGCTTAATCATGCTATTTCTTGTCTGATGATGAACAGAGAACTGGCAACCCAGATTC 1877
 QY 81 HisIleTyrLeuAspLysGluThrGlyLysProLysGlyAspAlaThrValSerTyrGlu 1900
 DB 1878 CACATCTACCTGGACAAAGAAACAGGAAAGCCCAAGCGAGTCCACAGTCTCTATGA 1937
 QY 101 AspProProThrAlaLysAlaAlaValGluTyrPheAsp----- 113
 DB 1938 GACCCACCACTGCGCAAGGCTGCGGTGAATGGTTGATGGTGAATGATCACTGCGCA 1997
 QY 114 -----GlyLysAspPheGlnGlySer 120
 DB 1998 TTCTTATCTAGTAATGATGATGTTCTGTGTTCTGTTCCAGGAAAGATTTTCAAGGAGC 2057
 QY 121 LysLeuLysValSerLeuAlaArgLysLysProPheMetAsnSerMetArgGlyLysLeu 140
 DB 2058 AAACCTTAAAGTCTCCCTGCTCGAAGAGACCTCCAAATGAACAGTATGGGGGTGCTCG 2117
 QY 141 ProProArgGluGlyLysGlyLysMetProProProLeuAspGly----- 154
 DB 2118 CCACCCCGTGAAGGAG 2177
 QY 155 -----GlyProGlyGlyProGly 160
 DB 2178 CTATGTCATTAATTTGCTGTTCTTCTGTTCTGTTCTGTTGATGATCAAGAGGCCAGGA 2237
 QY 161 GlyProGlyGlyProMetGlyLysMetGlyLysArgGlyLysAspArgGlyLysPhePro 180
 DB 2238 GGTCTGGGGAGAGCCCATGAGTGCATAGAGAGGCGGTGGAGAGATAGAGAGAGGCTTCCCT 2297
 QY 181 ProArgGlyProArgGlyLysSerArgLysAsnProSerGlyGlyLysAsnValGlnHisArg 200
 DB 2298 CCAG 2257
 QY 201 AlaGlyAspTyrGlnCysProAspPro----- 209
 DB 2358 GCTGAGAGCTGGAGAGTCTCCATCCGTATGACTGTCTGGGAAATGATACCTGAT 2417
 QY 210 -----GlyCysGlyAsnGlnAsnPheAlaTyrPheGln 220
 DB 2418 TTTCTGCTGATGATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2477
 QY 221 GlnCysAsnGlnCysLysAspLysProLysProGlyLysPheLeuProProPheProPro 240
 DB 2478 GAGTGCACCAAGTGAAGGCCCAAGAGGCTTCTCTCCGACACCTTCCGCC 2537
 QY 241 Pro----- 241
 DB 2538 CCGGATGAGTGCAGTTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2597
 QY 242 ---GlyLysAspArgGlyLysArgGlyLysProGlyLysMetArgGlyLysArgGlyLysLeu 260
 DB 2598 TTAGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2657
 QY 261 MetAspArgGlyLysProGlyLysMetPheArgGlyLysArgGlyLysLysAspArgGlyLys 280
 DB 2658 ATGATATCGTGGTGTCCCGTGAATGTTCAAGGTGGCCGCTGTGAAGCAAGAGTGGC 2717
 QY 281 PheArgGlyLysArgGlyLysMetAspArgGlyLysPheGlyLysLysArgArgGlyLysPro 300
 DB 2718 TTCCGTGGTGGCGGGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2777
 QY 301 GlyLysProProGlyLysProLeuMetGluGlnMetGlyLysArgArgGlyLysArgGlyLys 320

```
Db      2778 GGGGGGCCCCCTTGACCTTTGATGAAACAGATGGGAGGAAGAAGAGGAGGACGTGAGGA 2837
Qy      321 ProGlyLysMetAspLys----- 326
Db      2838 CCTGGAAAAATGGATTAAGTCTGTGAAAAAGCAGCTGTGGCCTTAACCGAAGG 2897
Qy      327 -----GlyLysHisArgGlnLysArgAspArgProTyr 338
Db      2898 CCTCTTTACCTTGCAAGGCGAGCACCGTCAGAGCGCAAGATCGGCCCTAC 2951
```

Search completed: February 20, 2005, 23:10:21
Job time : 451.771 secs

THIS Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - nucleic search, using frame_plus.p2n model

Run on: February 20, 2005, 22:00:21, Search time 149.958 Seconds
(without alignments)
3688.116 Million cell updates/sec

Title: US-10-791-017a-2_COPY_319_656
Perfect score: 1922
Sequence: 1 GAGGCGMAGSAGRGCGFKPKPG.....GGPGKDKGKHROGRRDRPY 338

Scoring table:
BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 1202784 seqs, 818138359 residues
Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODE=frame+ p2n.model -DEV=xlp
-O=/cgn2_1/USPTO.spool.p/US10791017/runat.17022005.125808.22107/app.query.fasta_1.1358
-Database=Patents.NA -QFMT=faastap -SUFFIX=p2n.rnt -MINMATCH=0.1 -LOOPT=0
-LOOEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blomsu62 -TRANS=human40.cdt
-LIST=45 -DOCLIGN=200 -THR SCORE=epct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US10791017@cgn2_1.141@runat.17022005.125808.22107 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -MAIN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database:

Issued_Patents_NA.*
1: /cgn2_6/ptodata/1/1na/5A.COMB.seq.*
2: /cgn2_6/ptodata/1/1na/5B.COMB.seq.*
3: /cgn2_6/ptodata/1/1na/6A.COMB.seq.*
4: /cgn2_6/ptodata/1/1na/6B.COMB.seq.*
5: /cgn2_6/ptodata/1/1na/PTCUS.COMB.seq.*
6: /cgn2_6/ptodata/1/1na/backfile1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1922	100.0	2371	2	US-08-343-443B-1
2	1810	94.2	1785	4	US-09-949-016-5043
3	1810	94.2	1785	4	US-09-949-016-5044
4	1632	84.9	6002	4	US-09-949-016-13696
5	1541	80.2	1783	4	US-09-949-016-1954
6	958	49.8	35784	4	US-09-949-016-16785
7	958	49.8	35784	4	US-09-949-016-16786
8	878	45.7	1939	4	US-09-919-039-322
9	609	31.7	411	4	US-09-621-976-13361
10	405	21.1	601	4	US-09-949-016-176641
11	405	21.1	601	4	US-09-949-016-17664
12	344	17.9	450	3	US-09-370-838-145

13	344	17.9	450	4	US-09-854-133-145	Sequence 145, App
14	289	15.0	1926	3	US-09-249-585A-4	Sequence 4, Appl1
15	289	15.0	1931	2	US-09-130-114-2	Sequence 2, Appl1
16	285	14.8	3147	2	US-08-781-802-7	Sequence 7, Appl1
17	285	14.8	3147	3	US-08-694-078-7	Sequence 7, Appl1
18	285	14.8	3147	3	US-09-058-260-7	Sequence 7, Appl1
19	275	14.3	2082	4	US-09-818-780-7	Sequence 7, Appl1
20	272	14.2	4551	4	US-09-949-016-2338	Sequence 2338, Ap
21	269.5	14.0	1558	4	US-09-949-016-2635	Sequence 2635, Ap
22	268.5	14.0	4167	4	US-09-169-768-12	Sequence 12, Appl1
23	267	13.9	6109	4	US-09-795-061-1	Sequence 1, Appl1
24	266.5	13.9	4568	4	US-09-949-016-2498	Sequence 2498, Ap
25	265	13.8	5387	4	US-09-949-016-265	Sequence 265, App
26	265	13.8	440365	3	US-09-103-840A-2	Sequence 2, Appl1
27	263.5	13.7	1771	2	US-08-533-692A-7	Sequence 7, Appl1
28	263.5	13.7	1771	2	US-08-511-872-1	Sequence 1, Appl1
29	263.5	13.7	1771	3	US-09-183-861-7	Sequence 7, Appl1
30	263.5	13.7	1771	3	US-09-022-765-7	Sequence 7, Appl1
31	263.5	13.7	1771	4	US-09-551-974A-7	Sequence 7, Appl1
32	263.5	13.7	1771	4	US-09-565-501A-7	Sequence 7, Appl1
33	263.5	13.7	1771	4	US-09-639-206A-7	Sequence 7, Appl1
34	263.5	13.7	1771	4	US-09-874-923-7	Sequence 7, Appl1
35	263.5	13.7	1771	4	US-08-798-841-7	Sequence 7, Appl1
36	263.5	13.7	3170	4	US-09-169-768-1	Sequence 1, Appl1
37	263.5	13.7	3171	4	US-09-169-768-15	Sequence 15, Appl1
38	263.5	13.7	3171	4	US-09-169-768-19	Sequence 19, Appl1
39	263.5	13.7	3349	4	US-09-169-768-13	Sequence 13, Appl1
40	263.5	13.7	3531	4	US-09-169-768-7	Sequence 7, Appl1
41	263.5	13.7	3541	4	US-09-169-768-9	Sequence 9, Appl1
42	263.5	13.7	4409	4	US-09-331-347C-22	Sequence 22, Appl1
43	263	13.7	6200	4	US-09-795-061-3	Sequence 3, Appl1
44	262.5	13.7	22218	4	US-09-949-016-14240	Sequence 14240, A
45	261.5	13.6	2070	4	US-09-949-016-5405	Sequence 5405, Ap

ALIGNMENTS

RESULT 1
US-08-343-443B-1
Sequence 1, Application US/08343443B
Patent No. 5968734
GENERAL INFORMATION:
APPLICANT: Aurias, Alain
APPLICANT: Delattre, Olivier
APPLICANT: Desmaze, Chantal
APPLICANT: Melot, Thomas
APPLICANT: Peter, Martine
APPLICANT: Ploougastel, Beatrice
APPLICANT: Thomas, Jessica
TITLE OF INVENTION: NUCLEIC ACID CORRESPONDING TO A GENE OF
TITLE OF INVENTION: CHROMOSOME 22 INVOLVED IN RECURRENT CHROMOSOMAL
TITLE OF INVENTION: TRANSLOCATIONS ASSOCIATED WITH THE DEVELOPMENT OF CANCEROUS
TITLE OF INVENTION: TUMORS AND NUCLEIC ACIDS OF FUSION RESULTING FROM SAID
TITLE OF INVENTION: TRANSLOCATIONS
NUMBER OF SEQUENCES: 129
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Weiser & Associates
STREET: 230 South Fifteenth Street
CITY: Philadelphia
STATE: PA
COUNTRY: USA
ZIP: 19102
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: AEDIT 1.0 DOS text editor
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/343,443B
FILING DATE: 18-NOV-1994
CLASSIFICATION: 514
PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT/FR93/00494
 FILING DATE: 19-MAY-1993
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: FR 92/06123
 FILING DATE: 20-MAY-1992
 ATTORNEY/AGENT INFORMATION:
 NAME: Weisner, Gerard J.
 REGISTRATION NUMBER: 19,763
 REFERENCE/DOCKET NUMBER: 989,6121P
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 215-875-8383
 TELEFAX: 215-875-8394
 INFORMATION FOR SEQ ID NO: 1:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 2371 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear
 FEATURE:
 NAME/KEY: CDS
 LOCATION: 25..1992
 US-08-343-443B-1

Alignment Scores:
 Pred. No.: 3,066-97 Length: 2371
 Score: 1922.00 Matches: 338
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 Gaps: 0

US-10-791-017a-2_COPY_319_656 (1-338) x US-08-343-443B-1 (1-2371)

QY 1 G1YGLVARG1YGLYMETG1YSER1AG1YGLUARG1YGLYPHEANLYSPROGLY 20
 DB 979 GAGAGACGCGGTGAAATGGGACGCTGAGAGCGAGTGGCTTCAATAGCCGTGGGA 1038
 QY 21 PROMETASPG1UG1YPROASPLEUASPLEU1YPROPROVALASPPROASPG1UASPSER 40
 DB 1039 CCCATGATGAGGACCAAGATCTGATGAGCCCTCTGATGATCAAGATGAACTCT 1098
 QY 41 ASPASER1A1IET1YR1VAL1G1NG1YLEUASAPSER1A1TH1LEUASAPLEU1A 60
 DB 1099 GACAAAGTGAATTATTTATGTAAGATTAATGACGTGATCTAGATGATCTGGCA 1158
 QY 61 ASPHEPHELYSG1NCYSG1YVAL1VAL1YMETASNL1YAR1GTH1YGL1N1PROMET1LE 80
 DB 1159 GACTTCTTTAAGCAGTGTGGGTGTGTTAAGATGAACAAGAACTGGGCAACCATGATC 1218
 QY 81 H1A1I1E1YR1LEUASPL1YSG1U1TH1YGL1YSPRO1YSG1YASPA1ATH1R1VAL1SER1YR1GLU 100
 DB 1219 CACATCTACCTGGACAAAGAAAGAAAGAAAGCCCAAGGCGATGACACAGTCTTATGAA 1278
 QY 101 ASPPROTH1R1A1IET1YR1VAL1A1A1VAL1G1UT1R1PHEASPG1YLYASPPHEG1NG1YSER 120
 DB 1279 GACCAACCCATGCGCAAGCGCTGCGTGAATGTTTATGGGAAAGATTTTCAAGGGAGC 1338
 QY 121 LYSLEULYVAL1SER1LEU1A1R1YGL1YSPRO1YMETASNL1YAR1GTH1YGL1YGL1YLEU 140
 DB 1339 AAACCTTAAAGTCTCCCTTCTCGGAGAAAGCTTCAATGAACAGTATGCGGGGTGCTCTG 1398
 QY 141 PROPROASPG1UG1YARG1YMET1YSPROPROASPLEU1YPROGLY1YPROGLY 160
 DB 1399 CCACCCCGTGAAGGAGAGGAGATGACACACACTCCGTGAGGTCCAGGAGGCGCCAGGA 1458
 QY 161 G1YPROGLY1YPROMETG1YARGMETG1YGL1YARG1YGL1YASPARG1YGL1YPHEPRO 180
 DB 1459 GGTCTGGGGAGCCCATGGGTGCGATGGAGGCGGTGAGAGATTAAGAGGCTTCCCT 1518
 QY 181 PROARG1YPROARG1YSE1YARG1YASNP1ROSE1YGL1YGL1YASNL1YGL1YH1A1R1G 200
 DB 1519 CCAAGAGGACCCCGGGGTTCCTCGAGGAGAAACCTCTGAGAGAGAAACGTCAACACCGA 1578

QY 201 A1AG1YASPT1R1G1NCYSPROASNP1ROGLYCYSG1YASNL1YASNL1A1TR1PARG1THR 220
 DB 1579 GCTGAGAGCTGCAAGTGTCCCAATCCGGTTCTGAAACCAACTCCCTGGAGAA 1638
 QY 221 GIUCYASNL1NCYEL1YAL1A1P1ROLYSPROGL1Y1PHELEUPROPROPH1P1RO 240
 DB 1639 GAGTGCAACCAATGTAAGGCCCAAGCTGAAAGCTTCTCCCGCACCTTCCGCC 1698
 QY 241 PROGL1Y1YASPARG1YARG1YGL1YPROGL1Y1YMET1YARG1YGL1YARG1YGL1YLEU 260
 DB 1699 CCGGGGTGATGTCGACAGAGTGGCCCTGATGATGCGGAGAGAGAGAGTGGCTTC 1758
 QY 261 METASPARG1YGL1YPROGL1Y1YMET1YMET1YARG1YGL1YARG1YGL1YASPARG1YGL1Y 280
 DB 1759 ATGATCGTGTGATGTCGCGGTGATGATGTTCAAGGTGGCCGATGATGACAGAGTGGC 1818
 QY 281 PHEARG1YGL1YARG1YMETASPARG1YGL1Y1YMET1YGL1Y1YARG1YGL1Y1YPRO 300
 DB 1819 TTCCTGTGTGGCCGGGCGATGACCGAGGTGGCTTGTGTGTGAGAAAGAGAGTGGCCCT 1878
 QY 301 G1YGLYPROPROGL1YPROLEUMETG1UG1METG1YGL1YARG1YGL1YARG1YGL1Y 320
 DB 1879 GGGGGGCCCCCTGAGCTTGTGATGAAACAGATGGGAGAGAAAGAGAGACGTGAGCA 1938
 QY 321 PROGL1Y1YMETASPL1YSG1YGL1Y1YARG1YGL1YARG1YGL1YARG1YGL1Y 338
 DB 1939 CCTGAGAAATGATTAAGCCAGCACCGTCAGAGCGCAGAGATGGGCCCTAC 1992

RESULT 2
 US-09-949-016-5043
 Sequence 5043, Application US/09949016
 Patent No. 6812339
 GENERAL INFORMATION:
 APPLICANT: VERTER, J. Craig et al.
 TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
 WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
 FILE REFERENCE: CL001307
 CURRENT APPLICATION NUMBER: US/09/949,016
 PRIOR FILING DATE: 2000-04-14
 PRIOR APPLICATION NUMBER: 60/241,755
 PRIOR FILING DATE: 2000-10-20
 PRIOR APPLICATION NUMBER: 60/237,768
 PRIOR FILING DATE: 2000-10-03
 PRIOR APPLICATION NUMBER: 60/231,498
 PRIOR FILING DATE: 2000-09-08
 NUMBER OF SEQ ID NOS: 207012
 SOFTWARE: FastSeq for Windows Version 4.0
 SEQ ID NO 5043
 LENGTH: 1785
 TYPE: DNA
 ORGANISM: Human
 US-09-949-016-5043

Alignment Scores:
 Pred. No.: 3,256-91 Length: 1785
 Score: 1810.00 Matches: 318
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 94.17% Indels: 0
 Gaps: 0

US-10-791-017a-2_COPY_319_656 (1-338) x US-09-949-016-5043 (1-1785)

QY 21 PROMETASPG1UG1YPROASPLEUASPLEU1YPROPROVALASPPROASPG1UASPSER 40
 DB 829 CCCATGATGAGGACCAAGATCTTGTATCTAGGCCCACTGTATATCAGATGAAGCTCT 888
 QY 41 ASPASER1A1IET1YR1VAL1G1NG1YLEUASAPSER1A1TH1LEUASAPLEU1A 60
 DB 889 GACAAAGTGAATTATTTATGTAAGATTAATGACAGTGTGATCTTATGATCTGGCA 948
 QY 61 ASPHEPHELYSG1NCYSG1YVAL1VAL1YMETASNL1YAR1GTH1YGL1N1PROMET1LE 80

```
Db      949 GACTCTTTAAGCAGTGGGGTGTGTAAGTGAACAAGAACTGGGCAACCATGATC 1008
Qy      81 HslltYrleuAaplysglUthrgLYLpProlysgLYAspAlaThrValSerTYrglu 100
Db      1009 CACATCTACTGACAGCAAGAAAGCAAGAAAGCCAAAGCCATGCCACAGTCTCTATGAA 1068
Qy      101 AspProThrAlaIysAlaIValGluTrpPheaspGLyYsaspPheGlnGlySer 120
Db      1069 GACCCACCCACTCCAAAGGCTGCGGTAATGTTGATGGAAAGATTTTCAAGGGAGC 1128
Qy      121 LysLeuLysValSerleuAlaArgLYLpProPheMetAsnSerMetArgLYLpLeu 140
Db      1129 AAACCTTAAGTCTCCCTGTCTCGGAAGAACCCTTCATGAACGTATGCGGGGTGGCTG 1188
Qy      141 ProProArgGLyLpArgLYLpMetProProPheLeuArgLYLpProGLYpProGLY 160
Db      1189 CCACCCCTGAGGGGAGAGGCAATGCCACCACTCCGTGAGATCCAGAGAGGCCCAAGA 1248
Qy      161 GLYpProGLYpProMetGLYpArgMetGLYLYpArgLYLpYAspArgGLYpPhePro 180
Db      1249 GGTCTCGGGGAGCCCATGGGTCCCATGGGAGGCCGTGGAGAGATAGAGAGGCTTCTC 1308
Qy      181 ProArgGLYpProArgLYSerArgLYAspProSerGLYLYpLYAsnValGlnHisArg 200
Db      1309 CCAAGAGAACCCCGGGGTTCCCAAGGAAACCCCTCTGAGAGAGAAACCTCCAGCAG 1368
Qy      201 AlaGLYAspTrpGlnCysProAspProGLYCYsGLYAsnGlnAsnPheAlaTrpArgThr 220
Db      1369 GCTGGAGACTGAGCAGTGTCCCAATCCGGGTTGGAGAAACAGAACTTCGCTGAGAGA 1428
Qy      221 GluCYAsnGlnCysLYsAlaProLYsProGLYpPheLeuProProPheProPro 240
Db      1429 GAGTGAACCAAGTGAAGGCCCAAGCCCTGAAGGCTTCTCCGCAACCTTTCGCGCC 1488
Qy      241 ProGLYLYAspArgGLYArgGLYpProGLYLYpMetArgGLYLYpArgGLYLYpLeu 260
Db      1489 CCGGGTGTGATCTGTGGCAGAGGTGGCTGTGTGGCATGGGGAGAGAAAGGTGGCTC 1548
Qy      261 MetAspArgGLYLYpProGLYLYpMetPheArgGLYLYpArgGLYLYpAspArgGLY 280
Db      1549 ATGATCGTGTGTGTCCCGGTGAATGTTCAAGAGTGGCCGTGTGGAGACAGAGGTGC 1608
Qy      281 PheArgGLYLYpArgGLYLYpMetAspArgGLYLYpPheGLYLYpArgGLYLYpPro 300
Db      1609 TTCGGTGTGGCCGAGGAGCATGAGCCGAGGTGGCTTGTGTGAGAGAAAGCATGGTGGCT 1668
Qy      301 GLYLYpProProGLYpProLeuMetGLYpGlnMetGLYLYpArgArgGLYLYpArgGLY 320
Db      1669 GGGGGGCCCCCTGACCTTGTATGAAACAGATGGAGAGAAAGAGAGAGAGAGAGAG 1728
Qy      321 ProGLYLYsMetAspLYsGLYpGlnHisArgGLYpGlnArgArgAspArgProTYr 338
Db      1729 CCGGAAAAAATGATTAAGCGAGACCGTCAGAGAGGAGAGATGGCCCTTAC 1782

RESULT 3
US-09-949-016-5044
; Sequence 5044, Application US/09949016
; Patent No. 681239
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: C1001307
; CURRENT APPLICATION NUMBER: US/09/949, 016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
```

```
; SEQ ID NO 5044
; LENGTH: 1785
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-5044

Alignment Scores:
Pred. No.: 3,25e-91 Length: 1785
Score: 1810.00 Matches: 318
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 94.17% Indels: 0
DB: 4 Gaps: 0

US-10-791-017a-2_COPY_319_656 (1-338) x US-09-949-016-5044 (1-1785)

Qy      21 ProMetAspGLyLpProAspLeuAapLeuGLYpProValAspProAspGLYAspSer 40
Db      829 CCCATGGATGAAGACCAAGATCTTGATCTTAGGCCCACTGTAGATCCAGATGAAGACTC 888
Qy      41 AspAsnSerAlaIleTYrValGlnGLYLeuAspAspSerValThrLeuAspAspLeuAla 60
Db      889 GACAAAGATGCAATTATATGACAAAGATTAAATGACAGTGTGACTTACATGATCTTGCA 948
Qy      61 AspPhePheLYsGlnCysGLYValLYsMetAsnLYsArgThrGLYpGlnProMetIle 80
Db      949 GACTCTTTAAGCAGTGGGGTGTGTAAGTGAACAAGAACTGGGCAACCATGATC 1008
Qy      81 HslltYrleuAaplysglUthrgLYLpProLYsGLYAspAlaThrValSerTYrglu 100
Db      1009 CACATCTACTGACAGCAAGAAAGCAAGAAAGCCAAAGCCATGCCACAGTCTCTATGAA 1068
Qy      101 AspProThrAlaIysAlaIValGluTrpPheaspGLYLYsaspPheGlnGlySer 120
Db      1069 GACCCACCCACTCCAAAGGCTGCGGTAATGTTGATGGAAAGATTTTCAAGGGAGC 1128
Qy      121 LysLeuLysValSerleuAlaArgLYLpProPheMetAsnSerMetArgGLYLYpLeu 140
Db      1129 AAACCTTAAGTCTCCCTGTCTCGGAAGAACCTTCATGAACGTATGCGGGGTGGCTG 1188
Qy      141 ProProArgGLYLYpArgLYLpMetProProPheLeuArgLYLpProGLYLYpProGLY 160
Db      1189 CCACCCCTGAGGGGAGAGGCAATGCCACCACTCCGTGAGATCCAGAGAGGCCCAAGA 1248
Qy      161 GLYpProGLYLYpProMetGLYpArgMetGLYLYpArgGLYLYpYAspArgGLYLYpPro 180
Db      1249 GGTCTCGGGGAGCCCATGGGTCCCATGGGAGGCCGTGGAGAGATAGAGAGGCTTCTC 1308
Qy      181 ProArgGLYpProArgLYSerArgLYAspProSerGLYLYpLYAsnValGlnHisArg 200
Db      1309 CCAAGAGAACCCCGGGGTTCCCAAGGAAACCCCTCTGAGAGAGAAACGTCCAGCAG 1368
Qy      201 AlaGLYAspTrpGlnCysProAspProGLYCYsGLYAsnGlnAsnPheAlaTrpArgThr 220
Db      1369 GCTGGAGACTGAGCAGTGTCCCAATCCGGGTTGGAGAAACCAAGATTCCTCGTGAAGAA 1428
Qy      241 ProGLYLYAspArgGLYArgGLYpProGLYLYpMetArgGLYLYpArgGLYLYpLeu 260
Db      1489 CCGGGTGTGATCTGTGGCAGAGGTGGCTGTGTGGCATGGGGAGAGAGAGGTGGCTC 1548
Qy      261 MetAspArgGLYLYpProGLYLYpMetPheArgGLYLYpArgGLYLYpYAspArgGLY 280
Db      1549 ATGATCGTGTGTGTCCCGGTGAATGTTCAAGAGTGGCCGTGTGGAGACAGAGGTGC 1608
Qy      281 PheArgGLYLYpArgGLYLYpMetAspArgGLYLYpPheGLYLYpArgArgGLYLYpPro 300
Db      1609 TTCGGTGTGGCCGAGGAGCATGAGCCGAGGTGGCTTGTGTGAGAGAAAGCATGGTGGCT 1668
Qy      301 GLYLYpProProGLYpProLeuMetGLYpGlnMetGLYLYpArgArgGLYLYpArgGLY 320
```

DB	Accession	Score	Length	Matches	Conservative	Mismatches	Indels	Gaps
Db	1669 GGGGGCCCCCTTGACCTTTATGGAACAATGGAGGAGGAAGAGGAGGACTGTGAGGA	1728						
QY	321 ProGlyLysMetAspLysGlyGluHisArgGlnGluArgAspArgProTyr	338						
Db	1729 CCTGGAAAAATGATATAAGGCGAGCACCTGTCAAGAGCGCAGAGATCGGCCCTTAC	1782						
RESULT 4								
US-09-949-016-13696	Application US/09949016							
Sequence 13696	Patent No. 6812339							
GENERAL INFORMATION:	APPLICANT: VENTER, J. Craig et al.							
TITLE OF INVENTION:	POLYMORPHISMS IN KNOWN GENES ASSOCIATED WITH HUMAN DISEASE, METHODS OF DETECTION AND USE THEREOF							
FILE REFERENCE:	CURRENT APPLICATION NUMBER: US/09/949,016							
CURRENT FILING DATE:	2000-04-14							
PRIOR APPLICATION NUMBER:	60/241,755							
PRIOR FILING DATE:	2000-10-20							
PRIOR APPLICATION NUMBER:	60/237,768							
PRIOR FILING DATE:	2000-10-03							
PRIOR APPLICATION NUMBER:	60/231,498							
PRIOR FILING DATE:	2000-09-08							
NUMBER OF SEQ ID NOS:	207012							
SOFTWARE:	FastSeq for Windows Version 4.0							
SEQ ID NO 13696								
LENGTH:	6002							
TYPE:	DNA							
ORGANISM:	Human							
US-09-949-016-13696								
Alignment Scores:								
Pred. No.:	6,14e-81	Length:	6002					
Score:	1632.00	Matches:	299					
Percent Similarity:	91.72%	Conservative:	11					
Best Local Similarity:	88.46%	Mismatches:	28					
Query Match:	84.91%	Indels:	2					
DB:	4	Gaps:	0					
US-10-791-017A-2_COPY 319_656 (1-338) x US-09-949-016-13696 (1-6002)								
QY	1 G1yG1yArG1yG1yMeG1ySeRa1aG1yG1yArG1yG1yPheAsnLysProG1yG1y	20						
Db	2988 GGAGGATGGCGTGAATGCGGACCGCTGGAGAGCAAGTTGGCTTCAATAAGCTGTGGA	3047						
QY	21 ProMetAspGlnGlyProAspLeuAspLeuGlyProProValAspProAspGluAspSer	40						
Db	3048 CCCATGAGTGAAGAGACCAATCTTGATCTTAAGCCCACTGTATATCCAGTGAAGACTCT	3107						
QY	41 AspAsnSerAlaIleTyrValGlnGlyLeuAsnAspSerValThrluAspAspLeuAla	60						
Db	3108 GACCAACAGTGCATTTATATGACAGATTTAAAGACAAATGTGACTTAAGTATCTGTGA	3167						
QY	61 AspPhePheLysGlnCysGlyValValLysMetAsnLysArgThrGlnProMetIle	80						
Db	3168 GACTCTTTAAGCAGGTGGGCTGTTAAGATGAACAAGAGACTGACCAACCATGATC	3227						
QY	81 HisIleTyrLeuAspLysGlnThrGlyLysProLysGlyAspAlaThrValSerTyrGlu	100						
Db	3228 CACACCTACCTGGACAGGAACAAGAAAGCCCAAGGTATATCCACAGTCTCTGTGA	3287						
QY	101 AspProProThrAlaLysAlaValAlaGluTrrPheAspGlyLysAspPheGlnGlySer	120						
Db	3288 GACTCACCCTACGTCACCAAGCTGCCGTGGATGTTTGAATGGAAAGATTTTCAAGGAGC	3347						
QY	121 LysLeuLysValSerLeuAlaArgLysPheProMetAsnSerMetArgGlyGlyLeu	140						
Db	3348 AAACCTTAAAGTCTCTCTGCTCGGAAGAGGCTCCACGTGAACAGTATGCAAGGGGTATG	3407						
QY	141 ProProArgGlnGlyArgLysMetProProLysLeuAspGlyGlyLysProGlyGlyProGly	160						
Db	3408 CCACCCCGTAGGGGCGAGGGAGTCCACCACTGCGCGAAGTCCAGAGGCGCCAGGA	3467						

QY	161	GLYPRGGLGVLGYPROMEGLYARMECEGLYGLYARAGGLYGLYVARSARGLYGLYLYPHEPRO	180
Db	3468	AGTCCTGGGGACCCCAATGGATGCATATGGAGAGCCCTGAGAGAGATGAGAGAGCCCTCCT	3527
QY	181	PROARGLYPROARGLYSERARGLYANPROSERGLYGLYVARNVALGHNHARG	200
Db	3528	CCAAAGAGACCCCAAGGGTTCCCGAGGAGAACCTCTGGAGAGGAGAAAGTCCAGACACCA	3587
QY	201	ALAAGLYAPRTPGINCYPROASNPROGLYCYSGLYAANGINANPHELAATTPARGTTR	220
Db	3588	GCTGAGACAGACGAGGTGCCCAATCCGGGGTGTGTGAAACCAAGACTTGCTGCGAGAGACA	3647
QY	221	GLUCYSAENGINCYSLYALAPROLYSPROGLINGLYPHELEUPROPROBHEPRO	240
Db	3648	GAGAGCAACAAAGTGTAGAGCTCCAAAGCTGAAGCTTCTCCGCCACCTTCCACACC	3707
QY	241	PROGLYGLYAPARAGLYVARGLYGLYLYPARGLYGLYMETARAGLYGLYVARGLYGLYLEU	260
Db	3708	CCGGGTGTGTATCATGCGAGAGGTGGCCCTGGTGCATGTGGGGAGAGAAAGAGTGGCTTC	3767
QY	261	METAPARAGLYVGLYPRPROGLYGLYMERPEARAGLYVGLYVARGLYGLYVAPARAGLYGLY	280
Db	3768	ATGATCATGTGTGTGTCCTGGGTGAATGTTCAGAGGTGGTGTGTGTAGACACGAAGTGGC	3827
QY	281	PEARAGLYGLYVARGLYMETAPARAGLYGLYPHEGLYGLYVARGLYGLYVARGLYGLYPRO	300
Db	3828	TTCCGTGTGGCTG-CGCAATGACCCGAGGTGGCTTTGGTGGAGAGAAACAGAGTGGCCCT	3886
QY	301	GLYLYP-PRO-PROGLYPROLEUMETGLINMETGLYVARGVARGLYVGLYVARGLYGLY	320
Db	3887	GGGGGGCCCCCGGA-CTTTGATGTMACCAATGGAGAGAAAGAGAGAGACGTGAAGGA	3945
QY	321	PROGLYLYSMETAPARAGLYGLYHNHARGINGLIUARGARGAPARPROTYR	338
Db	3946	CCTGGAATAACGATTAAGCGAGCACTGTCAAGAGCCGAGAGATCAGCCCTTAC	3999
RESULT 5			
US-09-949-016-1954			
Sequence 1954, Application US/09949016			
Patent No. 6812339			
GENERAL INFORMATION:			
APPLICANT: VENTER, J. Craig et al.			
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED			
TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF			
FILE REFERENCE: CL001307			
CURRENT APPLICATION NUMBER: US/09/949,016			
CURRENT FILING DATE: 2000-04-14			
PRIOR APPLICATION NUMBER: 60/241,755			
PRIOR FILING DATE: 2000-10-20			
PRIOR APPLICATION NUMBER: 60/237,768			
PRIOR FILING DATE: 2000-10-03			
PRIOR APPLICATION NUMBER: 60/231,498			
PRIOR FILING DATE: 2000-09-08			
NUMBER OF SEQ ID NOS: 207012			
SOFTWARE: FASTSEQ for Windows Version 4.0			
SEQ ID NO 1954			
LENGTH: 1783			
TYPE: DNA			
ORGANISM: Human			
US-09-949-016-1954			
Alignment Scores:			
Pred. No.: 1,876-76			
Score: 1541.00			
Percent Similarity: 91.82%			
Best Local Similarity: 88.68%			
Query Match: 80.18%			
DB: 4			
Gaps: 0			
US-10-791-017A-2_COPY_319_656 (1-338) x US-09-949-016-1954 (1-1783)			
21 ProMetApBgGLGVLGYPromEGLYARMECEGLYGLYARAGGLYGLYVARSARGLYGLYLYPHEPRO			

QY	154	-----	154
Db	32106	GGAGAAATTGTTGAATCTGGGGGGGTGGAGGTGCAATGACAAAGATGTCGCACTGCA	32155
QY	154	-----	154
Db	32166	CTCCAGCCTGGGCAACAGTGTGAGACTTCGTCCTCAAAAAAAAAAAAAAAAAATTTGTGGG	32225
QY	154	-----	154
Db	32226	AGCTCTGTTTCTGTAGAGCAGTGGAAACAGCTTCACAGGAGAGGGGGCTGATGCTCT	32285
QY	154	-----	154
Db	32286	GAGCCACACGGAAACACGGGACAGGTATGGGGAAATGACAGCATGTTCTGTGGGTTT	32345
QY	154	-----	154
Db	32346	TACTTATGATTTTATTTCTATAGCAAAATTTGGTGTCTACAGAGAAATGATTTGCTGTT	32405
QY	155	-----	155
Db	32406	TCCTGTGTTCTGTTGTGATGATCCAGAGAGGCCGAGAGTCTCGGGGAGCCCATGGGTGG	32465
QY	168	-----	168
Db	32466	CAATGGAGAGCGGTGAGAGATGATGAGAGGCTTCTCTCAAGAGAGACCCGGGGTTCCTCG	32525
QY	188	-----	188
Db	32526	AGGGAACCCCTCTGGAGGAGAAACGTCCAGCACCGAGCTGGAGACTGGCAGTGTCCAA	32585
QY	208	-----	208
Db	32586	TCCTGTA-TGTACTGTCTTGCGCAATTGATACCTTACAGATGAAAGCCACCTTCCCTCAC	32644
QY	210	-----	210
Db	32645	CCCATCCCACTTGAAGTGAATGCTGTCTGTCTTGAAGAAACAGATGATGACCTTATGG	32704
QY	210	-----	210
Db	32705	CTGTTAGGGAACATGATGACCCATTGACTGGACGCTTACAGAGCTTCTGAAGATTGATT	32764
QY	210	-----	210
Db	32765	TGACCTGTCTGTGGGTGCAATGCTGCTGAGAGCTGTGCTTAAAGCATGGGTGTACATA	32824
QY	210	-----	210
Db	32825	GATCTCTTGAATGAGTGTGTACCTGTTCACACACACCTTCTTGTATTCTTCTCT	32884
QY	211	-----	211
Db	32885	TAGTTCAATGGTGAATTTCTGCTGTGATGATTAATGTATGCAAGGGGTGTGGAAACAGAA	32944
QY	215	-----	215
Db	32945	CTTCGCTCGAGAAACAGAGTCAACAGTGTAAAGCCCAAGCTGTAAAGCTTCTCTCC	33004
QY	225	-----	225
Db	33005	GCCACCCCTTTCGCCCCCGGGGTAGTGCAGGTTTCATGATGTCCCTCAGCTTCTGTGT	33064
QY	241	-----	241
Db	33065	GCTAAACCTCTTTTCTTATTTGTGGGCTTGGTAAACAGAGTGCCTCTGCTTAAACAC	33124
QY	241	-----	241
Db	33125	TTTGAGTTGTGCTGTCTCATTTTCTAAATGTGACGCCGATGCCGAGATTGAGTGAAGTG	33184
QY	241	-----	241

```

Db      33185 TCTGCTTTGTTCTGCTGTGAGAAAGAAAGACAGACAGTCTTCCACAGTGTCCACAGGGCC 33244
Qy      242 -----GlyGlyAspArgGly 246
Db      33245 TCTGCAGCACCACCACTGACTGCTTGGCCCTTCATTTCTCACTTAATGGTGGTGTGTG 33304
Qy      246 YArgGlyGlyProGlyGlyMetArgGlyGlyYArgGlyGlyLeuMetAspArgGlyGlyPro 266
Db      33305 CAGAGTGGCCCTGCTGGTGCATGCGGGGAGAAAGAGTGGCCCTCATGGATCGTGTGTGTC 33364
Qy      266 OGlyGlyMetPheArgGlyGlyYArgGlyGlyAspArgGlyGlyPheArgGlyGlyYArgGly 286
Db      33365 CGGTGGAATGTTCAAGAGGTGGCCCGGTGGAGACAGAGGTGGCTTCCGTGTGGCCGGGG 33424
Qy      286 YMetAspArgGlyGlyPheGlyGlyGlyYArgArgGlyGlyProGlyGlyProProGlyPro 306
Db      33425 CATGACCCGAAGTGGCTTTGGTGGAGGAAGACAGAGTGGCCCTTGGGGGGCCCTTGAGACC 33484
Qy      306 OLeuMetGlyGlyInMetGlyGlyYArgArgGlyGlyYArgGlyGlyProGlyGlyMetAspArg 326
Db      33485 TTGATGGAACAGATGGAGGAAGAAAGAGAGAGAGTGGAGACCTGGAAAAATGATTA 33544
Qy      326 B----- 326
Db      33545 -GTAAAGTGTGATGAAAAAGCAGCTGTGGGCGCCAGACACAGTAAGAGACAGCCCTTCC 33603
Qy      326 ----- 326
Db      33604 CAGCTGTGTGGCGCAAGTCTCATGTCTTACGAAAGCTTGTGATGTGTGGAGAGAG 33663
Qy      326 ----- 326
Db      33664 CCAGAGAGGGGACACTGGGGGCTTGAAAGGGCTTCTCAACCCCTTCCATCTTAACGA 33723
Qy      327 -----GlyGlyHisArgGlyGlyIleArgGlyGlyAspArgProGly 338
Db      33724 AGGGCCCTCTTACTTCACTTGCAGAGGGAGACACCGTCAGAGCGCAGATCGCCCTAC 33781

RESULT 7
US-09-949-016-16786
; Sequence 16786, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CLO01307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16786
; LENGTH: 35784
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-16786

Alignment Scores:
Pred. No.: 3,336-43 Length: 35784
Score: 958.00 Matches: 233
Percent Similarity: 29.75% Conservative: 5
Best Local Similarity: 29.12% Mismatches: 11
Query Match: 49.84% Indels: 554
DB: 4 Gaps: 6

US-10-791-017A-2_COPY_319_656 (1-338) x US-09-949-016-16786 (1-35784)

```


Db 33545 -GTAAGTGTGTGTAAGAAAGAGCTGTGGCCCGCCAGGACACATGAAGAGACAGCCCTTCC 33603
QY 326 ----- 326
Db 33604 CAGCTGTGTGGCGCAAGTCCTCATGTCTTAGAAGCTTGATGTATGTGTGGAGAG 33663
QY 326 ----- 326
Db 33664 CCAGAGAGGGGACCTGGGGGCTCTGAGAGGCTTCCCTCCATCCCTTCCATTTCTAACCA 33723
QY 327 -----GlycyluHisArgGlnuArgArgAspArgProTyr 338
Db 33724 AGGGCCCTCTTACCTTTCAGAGAGCGACCCCTCAGAGAGCGAGATGCGCCCTAC 33781

RESULT 8
US-09-919-039-322
Sequence 322, Application US/09919039
Patent No. 6727066
GENERAL INFORMATION:
APPLICANT: Kaber, Matthew R.
TITLE OF INVENTION: GENES EXPRESSED IN TREATED HUMAN C3A LIVER CELL CULTURES
FILE REFERENCE: PA-0035 US
CURRENT APPLICATION NUMBER: US/09/919,039
CURRENT FILING DATE: 2002-09-09
PRIOR FILING DATE: 2000-07-28
NUMBER OF SEQ ID NOS: 401
SOFTWARE: PERL Program
SEQ ID NO 322
LENGTH: 1939
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc feature
OTHER INFORMATION: Incyte ID No. 6727066 478620.53
US-09-919-039-322

Alignment Scores:
Pred. No.: 4,9e-40 Length: 1939
Score: 878.00 Matches: 182
Percent Similarity: 63.37% Conservative: 36
Best Local Similarity: 52.91% Mismatches: 54
Query Match: 45.68% Indels: 72
DB: 4 Gaps: 13

US-10-791-017a-2_COPY_319_656 (1-338) x US-09-919-039-322 (1-1939)

QY 1 G1G1YARGLYGLYMeTGLYSerAlaG1YUArgG1YGLYPheAsnLYSProG1YGLY 20
Db 825 GGAGGCGAGGTGGCATGGCCGGAAGT---GACCGTGTGGCTTCAATTAATTGGTGGC 881
QY 21 PrometApg1uG1YProApleuApg1YProProValAapProApg1uAapSer 40
Db 882 CCGGGGACCAAGATCACGTCATGAC-----TCCGAACAGCATTAATTC 926
QY 41 AspaAserAla1eTYrValG1ng1YleuAsnAapSerValThreleuAapleuAla 60
Db 927 GACAAACAACCACTTTGTGCAAGGCTGGGAGATTTTCATTAATGAGTGTGGCT 966
QY 61 AspPhePhe1yG1nCYsG1YVal1ValYMeTAsnLYSArgThrg1YGLInProMetile 80
Db 987 GATTACTTCAAGCATGTGATATTATTTAAGCAAAACAGAAACGGGACGCCCATGATT 1046
QY 81 He1leTYrleuApg1YSerG1YUArgG1YLeuProLYSArgAlaThrValSerTYrG1u 100
Db 1047 AATTGTACACAGACAGGAGAACTGGCAAGCTGAAGGAGAGCAACGGCTCTTTGAT 1106
QY 101 AspProProThrAla1ySAla1aValG1UTrPheAsnG1YLYSAspPheG1ng1YSer 120
Db 1107 GACCCACCTTACCTAAAGCAGCTATTGACTGTGTTGATGTAAAGAAATTCCTCCGGAAT 1166
QY 121 LYleuLYSValSerleuAlaArgLYSProProMetAapSerMeTArgLYG1YLeu 140

Db 1167 CCTATCAAGGTCTCATTTTGTCTACTGCGCGGAGACTTAAT-----CGGGGTGTGGC 1220
QY 141 ProProApg1uG1YARGLYMeTProProProleuArgLYGLYProG1YLYProG1Y 160
Db 1221 AATGTCTGGAGGCGCGAGGG-----CGAGAGAGACCCATGGCCCTGTGA 1265
QY 161 G1YProG1YGLYProMeTGLYArgG1YGLYAspArgLYGLYPhePro 180
Db 1266 GGCCTATGAGGTGTG-----GGCACTGTGTGTGTGGCCGAGAGGATTTCC 1313
QY 181 ProArgG1YProArgG1YSerArgLYSArgProSerG1YGLYAsnValGln1SArg 200
Db 1314 AGTGAAGT-----GATGGCGGTGAAGACAGCAGCA 1346
QY 201 AlaG1YAspTYrG1nCYsProAapSerProG1YCYsG1YAsnG1nAsnPheAlaThrPAGTTr 220
Db 1347 GCTGTGACTGGAAGTGAAGTCTTAATCCACCTGTGAGAAATGAACTTCTTGGAGGAAT 1406
QY 221 G1UCYAsnG1nCYsLYSAlaProLYSProG1uG1YPhleuProProPheProPro 240
Db 1407 GAATGCAACCAAGTGAAGGCCCCCTTAACCAATGGC----- 1442
QY 241 ProG1YGLYAspArgG1YArgLYGLYProG1YLYMeTArg---G1YGLYArgG1YGLY 259
Db 1443 CCAGGA-----GGGGGACCAAGTGGCTTCACATGGGGGTAACTACGGG 1487
QY 260 LeuMetAapArgG1YGLYProG1YLYMeTPhleuArgLY-----G1YArgG1YGLY 276
Db 1488 GATGATCGTCGTGGTGGCGAGAGAGCTGTGATCGAGGGGCTACCGGGCCGCGGGCGG 1547
QY 277 AspArgG1YGLYPheArgLYGLYArg---G1YMeTAspArgG1YGLYPheG1YGLY 295
Db 1548 GACCGTGAAGGCTTCCGAGGGGCGGGGTGTGGGACAGAGTGGCTTT----- 1598
QY 296 ArgArgG1YLYProG1YLYProProG1YProleuMeTGLInMeTGLYArgArg 315
Db 1598 ----- 1598
QY 316 G1YGLYARGLYGLYProG1YLYMeTAsp---LYsG1YUHisArgG1ng1YUArgArg 334
Db 1599 -----GGCCCTGGCAAGATGATTCAGGGGTGAGACAGACAGATCCAGG 1646
QY 335 AspArgProTyr 338
Db 1647 GAGAGGCGGTAT 1658

RESULT 9
US-09-621-976-13361
Sequence 13361, Application US/09621976
Patent No. 6639063
GENERAL INFORMATION:
APPLICANT: Dumas Milne Edwards, J.B.
APPLICANT: Jobert, S.
APPLICANT: Giordano, J.Y.
TITLE OF INVENTION: ESTs and Encoded Human Proteins.
FILE REFERENCE: GENSET 054PR2
CURRENT APPLICATION NUMBER: US/09/621,976
NUMBER OF SEQ ID NOS: 19335
SOFTWARE: Patent.pm
SEQ ID NO 13361
LENGTH: 411
TYPE: DNA
ORGANISM: Homo sapiens
US-09-621-976-13361

Alignment Scores:
Pred. No.: 6,33e-26 Length: 411
Score: 609.00 Matches: 116
Percent Similarity: 99.15% Conservative: 0
Best Local Similarity: 99.15% Mismatches: 0
Query Match: 31.69% Indels: 1

```

DB: 4 Gaps: 0
US-10-791-017A-2_COPY_319_656 (1-338) x US-09-621-976-13361 (1-411)
QY 20 G1YPrometAsg1UG1yProaPleuAaspLeuG1yProProValaPProaPgiUasp 39
DB 62 GGAACCATGATGATGAAGGACCAAGATCTTGATCTTAGGCCCACTTAATCAATGATGAAGC 121
QY 40 SerPaPasnSerAlaIleTyValGing1yLeuAasnApsSerValThrLeuAaspLeu 59
DB 122 TCTGCAACAGTGCATATTATGTAACAAGATTAAATGACAGTGTGACTGATGATGTC 181
QY 60 AlaAapPhePheLeuG1yValValValyPMeAasnLybArghThrg1yG1nPromet 79
DB 182 GGAGACTCTTTTAAGAGAGTGGGGTGTGTTAAATATAACAAGAACTGGGCAACCAATG 241
QY 80 I1eHs1leTyLeuAapLyseG1uThrg1yLysProLyseG1yAspAlaThrValSerTy 99
DB 242 ATCCACATTTACTTGCAGCAAGAAACAGAAAGCCAAAGGCGATTGCCACATGTCTTA 301
QY 99 tG1uBspProProThrAlaLyseAlaValG1uTrPheAaspG1yLysAaspPheG1ng1 119
DB 302 TGAAGCCACCAAGCCAGTCCAGAGCTCCGAGAAATGTTGATGAGGAATTTTCAAG 361
QY 119 ySerLyseLeuLyValSerLeuAlaArgLySbLysPProPmetAasnSer 135
DB 362 GAGCAAACTTAAAGTCTCCCTTGCTCGAAGAAAGCTCCCAATGAACAGT 410

RESULT 10
US-09-949-016-176641
; Sequence 176641, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO: 176641
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-176641

Alignment Scores:
Pred. No.: 1,43e-14 Length: 601
Score: 405.00 Matches: 82
Percent Similarity: 54.67% Conservative: 0
Best Local Similarity: 54.67% Mismatches: 1
Query Match: 21.07% Indels: 68
Gaps: 1

US-10-791-017A-2_COPY_319_656 (1-338) x US-09-949-016-176641 (1-601)
QY 256 G1YArgG1yG1yLeuMetAaspArgG1yG1yProG1yG1yMetPheArgG1yG1yArgG1y 275
DB 2 GGAAGAGGTGGCTCATGATGATGTGTGTCCTCCGGGAGATGTCAGAGGTGGCCGTGT 61
QY 276 G1yAaspArgG1yG1yPheArgG1yG1yArgG1yMetAaspArgG1yG1yPheG1yG1y 295
DB 62 GGAAGCAAGAGGTGGCTTCCTGTGTGGCCGGGGCATGAGACGAAGTGGCTTTGTGTGAAGA 121
QY 296 ArgArgG1yG1yProG1yG1yProProG1yProLeuMetG1uG1uMetG1yG1yArgG1y 315

```

```

Db      122 AGACGAGGTGGCCCTTGCGGGGGCCCCCTTGACCTTTGATGAAACAGATGGAGAGAGAAGA 181
Qy      . . . 316 G1yG1yArG1yG1yProG1yLySmEaSpLys----- 326
Db      182 GGAGGACCGTGGAGGACCTGGAAAAATGATTA- GTAAGTGTGTGTGAAAAAGCAGCTGTGG 240
Qy      326 ----- 326
Db      241 GCCGCCAGGACAGTAAAGAGACAGACCCCTTCCAGCTGTGTGGCGCAAGTCTCATGTC 300
Qy      326 ----- 326
Db      301 KCTAGGAAGCTTGTGATGATGTGTTGGAGAGAGCCAGAAAGGGGACCTGGGGCTTGGA 360
Qy      327 -----g1yG1u 328
Db      361 AGGCGTTCCTACCCCTTCCACTTCTAAACGAAGGGCCCTCTTAACTTGCAGAGGCGAG 420
Qy      329 HisArG1InG1uArGArGAspArGProTyR 338
Db      421 CACCGTCAGAGCGCAGAGATGGCGCTTAC 450

RESULT 11
US-09-949-016-176664
; Sequence 176664, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CLO01107
; CURRENT APPLICATION NUMBER: US/09/949, 016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 176664
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-176664

Alignment Scores:
Pred. No.: 1.43e-14 Length: 601
Percent Similarity: 405.00 Matches: 82
Best Local Similarity: 54.67% Conservative: 0
Query Match: 21.07% Mismatches: 1
Db: 4 Gaps: 68

US-10-791-017A-2_COPY_319_656 (1-338) x US-09-949-016-176664 (1-601)
Qy      256 G1yArG1yG1yLyuMeTaSpArG1yG1yProG1yG1yMeTaPhArG1yG1yArG1y 275
Db      2 GGAAGAGGTGGCCCTCATGATGATGTGTGTGTCCCGGTGGAATTTAGAGGTGGCGGTGT 61
Qy      276 G1yAsPaRgG1yG1yPhArG1yG1yArG1yMeTaSpArG1yG1yPhG1yG1yG1y 295
Db      62 GGAACAGAGGTGGCTTCCGTGTGTGTCCCGGGGACAGACCGAGGTGGCTTTGTGTGAGGA 121
Qy      296 ArGArG1yG1yProG1yG1yProBrog1yProLyuMeTg1uG1uMeTg1yG1yArGAr 315
Db      122 AGACGAGGTGGCCCTGGGGGGCCCCCTTGACCTTTGATGAAACAGATGGAGAGAGAAGA 181
Qy      316 G1yG1yArG1yG1yProG1yLyMeTaSpLys----- 326
Db      182 GGAGGACCGTGGAGGACCTGGAAAAATGATTA- GTAAGTGTGTGTGAAAAAGCAGCTGTGG 240

```

```
QY 326 ----- 326
Db 241 GCCGCCAGCAGCAGTAGAGAGACGCCCTTCCAGCTTGCGCAAGTCTCATGTC 300
QY 326 ----- 326
Db 301 KCTAGAGACTGTGTATGATGTTGGAGAGAGACCAAGAGGACCTGGGGCTCTGGA 360
QY 327 ----- 327
Db 361 AGGGCTCTCTACCCCTTCCATCTTAACGAGGGCCCTTTACTTGCAGAGGCGAG 420
QY 329 H1ARGINGLUAARGAARGAPARPROLYR 338
Db 421 CACCGTCAGAGGCGCAGAGTGGCGCTTAC 450

RESULT 12
US-09-370-838-145
; Sequence 145, Application US/09370838
; Patent No. 6444425
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Mohamath, Radooh
; APPLICANT: Secrist, Heather
; TITLE OF INVENTION: COMPOUNDS FOR THERAPY AND DIAGNOSIS OF
; TITLE OF INVENTION: LUNG CANCER AND METHODS FOR THEIR USE
; FILE REFERENCE: 210121.475C1
; CURRENT APPLICATION NUMBER: US/09/370,838
; EARLIER FILING DATE: 1999-08-09
; EARLIER APPLICATION NUMBER: US 09/285,323
; NUMBER OF SEQ ID NOS: 289
; SOFTWARE: FaastSeq for Windows Version 3.0
; SEQ ID NO 145
; LENGTH: 450
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-370-838-145

Alignment Scores:
Pred. No.: 2,4e-11 Length: 450
Score: 344.00 Matches: 64
Percent Similarity: 79.468 Conservative: 25
Best Local Similarity: 57.148 Mismatches: 16
Query Match: 17.908 Indels: 7
Gaps: 2

US-10-791-017a-2_COPY_319_656 (1-338) x US-09-370-838-145 (1-450)
QY 1 G1G1YARG1G1YMETG1YSER1AG1YGLUARG1YGLYPHEN1LYSPROG1YGLY 20
Db 134 GGAGGAGAGGTGGCATGGCGGAGAGT--GACCGTGGTGGCTTCATTAATTGGTGGC 190
QY 21 ProMetAspGluGlyProAspLeuGlyProProValAspProAspGluAspSer 40
Db 191 CCTCGGACCAAGATCAGTCATGAC-----TCCGAACAGGATTAATCA 235
QY 41 AspAsnSerAlaIleTyValGlnGlyLeuAsnAspSerValThreLeuAspLeuAla 60
Db 236 GACAAACAACACATCTTGTGCAAGGCTGGTGAGAAATTTCAATTGAGCTGGGCT 295
QY 61 AspPhePheLysGlnCysGlyValValLysMetAsnLysArgThrGlyGlnProMetIle 80
Db 296 GATTACTTCAGAGAGATTGGTATATTATTAAGACAAACAAAGAAAACGGGACGCCCATGATT 355
QY 81 H1e1IeTyLeuAspLysGlnThrGlyLysProLysGlyAspAlaThrValSerTyGlu 100
Db 356 AATTGTACACAGACAGGAGAACTGGCAAGCTGAGAGGAGGACGACGCTCTTTGAT 415
QY 101 AspProProThrAlaLysAlaAla-ValGluTrp 111
Db 416 GACCCACCTTCAGCTAAAGCAGCCTATTGACTGG 449
```

```
RESULT 13
US-09-854-133-145
; Sequence 145, Application US/09854133
; Patent No. 6759508
; GENERAL INFORMATION:
; APPLICANT: Lodes, Michael J.
; APPLICANT: Mohamath, Radooh
; APPLICANT: Henderson, Robert A.
; APPLICANT: Benson, Darin R.
; APPLICANT: Secrist, Heather
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
; TITLE OF INVENTION: THE THERAPY AND DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.475C10
; CURRENT APPLICATION NUMBER: US/09/854,133
; CURRENT FILING DATE: 2001-05-11
; NUMBER OF SEQ ID NOS: 735
; SOFTWARE: FaastSeq for Windows Version 3.0
; SEQ ID NO 145
; LENGTH: 450
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-854-133-145

Alignment Scores:
Pred. No.: 2,4e-11 Length: 450
Score: 344.00 Matches: 64
Percent Similarity: 79.468 Conservative: 25
Best Local Similarity: 57.148 Mismatches: 16
Query Match: 17.908 Indels: 7
Gaps: 2

US-10-791-017a-2_COPY_319_656 (1-338) x US-09-854-133-145 (1-450)
QY 1 G1G1YARG1G1YMETG1YSER1AG1YGLUARG1YGLYPHEN1LYSPROG1YGLY 20
Db 134 GGAGGAGAGGTGGCATGGCGGAGAGT--GACCGTGGTGGCTTCATTAATTGGTGGC 190
QY 21 ProMetAspGluGlyProAspLeuGlyProProValAspProAspGluAspSer 40
Db 191 CCTCGGACCAAGATCAGTCATGAC-----TCCGAACAGGATTAATCA 235
QY 41 AspAsnSerAlaIleTyValGlnGlyLeuAsnAspSerValThreLeuAspLeuAla 60
Db 236 GACAAACAACACATCTTGTGCAAGGCTGGTGAGAAATTTCAATTGAGCTGGGCT 295
QY 61 AspPhePheLysGlnCysGlyValValLysMetAsnLysArgThrGlyGlnProMetIle 80
Db 296 GATTACTTCAGAGAGATTGGTATATTATTAAGACAAACAAAGAAAACGGGACGCCCATGATT 355
QY 81 H1e1IeTyLeuAspLysGlnThrGlyLysProLysGlyAspAlaThrValSerTyGlu 100
Db 356 AATTGTACACAGACAGGAGAACTGGCAAGCTGAGAGGAGGACGACGCTCTTTGAT 415
QY 101 AspProProThrAlaLysAlaAla-ValGluTrp 111
Db 416 GACCCACCTTCAGCTAAAGCAGCCTATTGACTGG 449

RESULT 14
US-09-249-585A-4/C
; Sequence 4, Application US/09249585A
; Patent No. 6417002
; GENERAL INFORMATION:
; APPLICANT: Hotlick, Robert
; APPLICANT: Mohamath, Radooh
; TITLE OF INVENTION: METHOD FOR MAINTENANCE AND SELECTION OF EPISOMES
; FILE REFERENCE: 0867/0D905
; CURRENT APPLICATION NUMBER: US/09/249,585A
; CURRENT FILING DATE: 1999-02-11
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 4
; LENGTH: 1926
; TYPE: DNA
```

ORGANISM: Epstein Barr Virus
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)..(1926)
OTHER INFORMATION: template strand of EBNA-1 DNA
US-09-249-585A-4

Alignment Scores:

Pred. No.:	1.02e-07	Length:	1926
Score:	289.00	Matches:	131
Percent Similarity:	40.15%	Conservative:	28
Best Local Similarity:	33.08%	Mismatches:	124
Query Match:	15.04%	Indels:	114
DB:	3	Gaps:	22

US-10-791-017a-2_COPY_319_656 (1-338) x US-09-249-585A-4 (1-1926)

```
QY 1 G1YGYARGLYGLYMETGLYSERIALGYLUARGLYGLYPHEANLYSPROGLY--- 19
DB 1388 GGAGGAAAAAAGCGGAGCGAGAGTAGTGAGAGTGGGCCCCAGGTCAAGAACCGGAGAGA 1329
QY 20 G1Y-PrometAapGLyGLYProAapLeuAapLeuGLYProPro-----Va 34
DB 1328 GGACCCAGTAGAC---GCCCGGGAGCGAGTAGCGAGGCGCCCGTCAGTCCAGTGT 1272
QY 34 LAAPPROAapGLYUapSerAapSerAa1a1eTyValGlnG1YLeuAapSerVa 54
DB 1271 AGACCCGCGTGAAGAACCA---TAAGTTTATTATGCGGAGGG----- 1229
QY 54 1ThLeuAapLeuAa1aapPhePheLyGlnCyG1YValValLyS----- 70
DB 1228 -----GATGTCCCACTTTTATCCGAGAGATGACCTCCCGGAGCGGCACCT 1179
QY 71 -----MetAenLySArgThG1YGLNProMet11e1a1eTy 83
DB 1178 CTGGGCTACTACTACTGATGATGACCTGAGAGACCGGAGAAAAGAGTGCAGTGG-- 1121
QY 83 rLeuAapLyGLYThrG1YLyS-----ProLySGLYAspAlaThrValSerTyGLUa 101
DB 1120 -CTGAGAGAGGGAGCGGAGAAAGTCTGAAAGGGGGAGCCGAGAAAGTGCAGAGATGG 1062
QY 101 pProProThrAlaLyS----- 106
DB 1061 GCCGCGGAGGTATGAGAGAGCTGGGCGGAGTATGAGAGAGCTGGGCGCGAGGT 1002
QY 107 -----AlaA1aValG1UTrpPheAapGLYLyAapPheGlnG1YSerTy 121
DB 1001 GACGAGAGAGCTGGGCGCGAGGTGAGAGAGAGAGCGGAGAGAGAGAGAGAGAGAG 942
QY 121 bLeuLyValSerLeuA1aArgLybLybProProMetAapSerMetAapGLYGLYLeuPr 141
DB 941 GGA-CGAGAGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 900
QY 141 oProAapGLYGLYArgGLYMetProProProLeuAapGLYGLYProGLY-----G1YPr 159
DB 899 -----CGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 859
QY 159 oG1YGLYProGLYGLYProMetGLYArgMetGLYGLYArgGLYGLYAspArgGLY----- 177
DB 858 GGGAGAGAGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 802
QY 178 -----G1YPhrProProAapGLYProAapGLYSerAa1a1eTyAspProSerGLYGLY 195
DB 801 AGGACGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 742
QY 195 yAenValGlnH1aArgAlaGLYAspTrGlnCybProAapProGLYCyG1YAsnGlna 215
DB 741 AGAGAGAGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 662
QY 215 nPheA1aTrpArgThrG1NcybAanG1NcybLybA1aProLybProGlnG1NcybLeuPr 235
DB 681 ACGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 669
```

```
QY 235 oProProPheProProProGLYGLYAspArgGLYArgGLYGLYProGLYGLYMetArgGLY 255
DB 668 -----GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 634
QY 255 YG1YArgGLYGLYLeuMetAapArgGLYGLYProGLYGLYMetPheArgGLYGLYArgGLY 275
DB 633 AGGACGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 574
QY 275 YG1Y-----AspArgGLYGLYPheArgGLYGLYArgGLYGLYMet----- 287
DB 573 AGGACGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 514
QY 288 -----AspArgGLYGLYPheArgGLYGLYArgGLYGLYMet-----ArgGLYGLYProGLYGLY 302
DB 513 AGGAGAGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 454
QY 302 YProProGLYProLeuMetGLYGLYMetGLYGLYArgGLYGLYArgGLYGLYArgGLYGLYPr 321
DB 453 AGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 406
QY 321 oG1Y-LysMetAapLyGLYGLYH1aArgGlnG1NcybArgAap 335
DB 405 GGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 362
```

RESULT 15

US-09-130-114-2/C
Sequence 2, Application US/09130114
Patent No. 5976807

GENERAL INFORMATION:

APPLICANT: Horlick, Robert A.
APPLICANT: Robbins, Alan K.

TITLE OF INVENTION: Eukaryotic Cells Stably Expressing Genes

FILE REFERENCE: 0867/1D903US1

CURRENT APPLICATION NUMBER: US/09/130, 114

NUMBER OF SEQ ID NOS: 36

SOFTWARE: FastSeq for Windows Version 3.0

SEQ ID NO 2

LENGTH: 1931

TYPE: DNA

ORGANISM: EBNA

US-09-130-114-2

Alignment Scores:

Pred. No.:	1.02e-07	Length:	1931
Score:	289.00	Matches:	131
Percent Similarity:	40.15%	Conservative:	28
Best Local Similarity:	33.08%	Mismatches:	124
Query Match:	15.04%	Indels:	114
DB:	2	Gaps:	22

US-10-791-017a-2_COPY_319_656 (1-338) x US-09-130-114-2 (1-1931)

```
QY 1 G1YGYARGLYGLYMETGLYSERIALGYLUARGLYGLYPHEANLYSPROGLY--- 19
DB 1388 GGAGGAAAAAAGCGGAGAGAGTAGTGAGAGTGGGCCCCAGGTCAAGAACCGGAGAGA 1329
QY 20 G1Y-PrometAapGLyGLYProAapLeuAapLeuGLYProPro-----Va 34
DB 1328 GGACCCAGTAGAC---GCCCGGGAGCGAGTAGCGAGGCGCCCGTCAGTCCAGTGT 1272
QY 34 LAAPPROAapGLYUapSerAapSerAa1a1eTyValGlnG1YLeuAapSerVa 54
DB 1271 AGACCCGCGTGAAGAACCA---TAAGTTTATTATGCGGAGGG----- 1229
QY 54 1ThLeuAapLeuAa1aapPhePheLyGlnCyG1YValValLyS----- 70
DB 1228 -----GATGTCCCACTTTTATCCGAGAGATGACCTCCCGGAGCGGCACCT 1179
QY 71 -----MetAenLySArgThrG1NcybAanG1NcybLybA1aProLybProGlnG1NcybLeuPr 83
```

Db 1178 CTGGGCTACTACTAGTATGACCTGAGAGACCGGAGAAAGAGTGCAGGTG-- 1121
QY rleuAapLyvgLutThrGlyLys-----ProLyvgLyAapAlaThrValSerTyrgLudA 101
|||:::|
Db 1120 -CTGGAGAGGGGAGCCGAGAAAGTCTGAAGGGGGGAGCCGAGAAAGTGCAGAGATGGG 1062
QY 101 pProProThrAlaLys----- 106
|||
Db 1061 GTCGCGGAGAGTATGAGAGAGCTGGGGCCGAGGTGATGAGAGAGCTGGGGCCGAGAGT 1002
QY 107 -----AlaAlaValGluTrpPheAapGlyLysAapPheGlnGlySerLy 121
|||:::|
Db 1001 GACGAGAGAGCTGGGGCCGAGAGTGCAGAGACGAGAGACGAGAGACGAGAGACGAGAGAG 942
QY 121 sLeuLyvValSerLeuAlaArgLyvLyvProProMetAapSerMetArgGlyGlyLeupr 141
: : :
Db 941 GGA-CGAGAGACGGGGAGAGACGGGGAG-----GAGCAGAGACGGGGAGAG----- 900
QY 141 oProArgGluGlyArgGlyMetProProProLeuArgGlyGlyProGly-----GlyPr 159
|||
Db 899 ----CGGGAGAGACGAGAG-----CGGGAGAGACGGGGAGAGACGAGAGACG 859
QY 159 oGlyGlyProGlyGlyProMetGlyArgMetGlyGlyArgGlyGlyAapArgGly----- 177
|||
Db 858 GGGAGAGACGGGGAGAGAGAGAGAGACGG-----GGAGAGACGGGGAGAGAGAGACGGGG 802
QY 178 -----GlyPheProProArgGlyProArgGlySerArgGlyAapProSerGlyGly 195
|||
Db 801 AGGACGGGGAGAGACGAGAGACGGGGAGAGACGGGGAGAGACGGGGAGAGACGGGGAGG 742
QY 195 yAapValGlnHisArgAlaArgLysArgTrpGlnCyvProAapProGlyCyvGlyAapGlnAa 215
|||
Db 741 ACAGAGACGGGGAGAGACGAGAGACGGGGAGAGACGAGAGACGGGGAGAGACGGGGAGG 682
QY 215 nPheAlaTrpArgThrGluCyvAapGlnCyvLyvAlaProLyvProGluGlyPheLeupr 235
|||
Db 681 ACCGGGAGAGACGG----- 669
QY 235 oProProPheProProProGlyGlyAapArgGlyArgGlyGlyProGlyGlyMetArgGly 255
|||
Db 668 -----GGAGAGAGACGAGAGACGGGGAGAGACGAGAGACGAGAGACGAGAGACG 634
QY 255 yGlyArgGlyGlyLeuMetAapArgGlyGlyProGlyGlyMetPheArgGlyGlyArgGly 275
|||
Db 633 AGGACGGGGAGAGACGGGGAGAGACGAGAGACGGGGAGAGACGAGAGACGAGAGACGAGG 574
QY 275 yGly-----AapArgGlyGlyPheArgGlyGlyArgGlyMet----- 287
|||
Db 573 AGGACGGGGAGG 514
QY 288 -----AapArgGlyGlyPheGlyGlyGlyArgGlyGlyProGlyGly 302
|||
Db 513 ACAG 454
QY 302 yProProGlyProLeuMetGluGlnMetGlyGlyArgGlyGlyArgGlyGlyArgGlyGlyPr 321
|||
Db 453 AGGACGAGAG-----CGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGC 406
QY 321 oGly-LysMetAapLyvgLyGluHisArgGlnGluArgArgAap 335
|||
Db 405 GGGAGC 362

Search completed: February 21, 2005, 04:23:33
Job time : 225.958 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - nucleic search, using frame_plus.p2n model

Run on: February 20, 2005, 21:44:36 ; Search time 2515 Seconds
(without alignments)
5075.236 Million cell updates/sec

Title: US-10-791-017a-2_COPY_319_656

Perfect score: 1922

Sequence: 1 GGRGCMGSGRGRGFKPG.....GGRGKDKGRHGRDRPRY 338

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-Processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODE=frame+ .p2n.model -DEV=xip
-Q/cg2.1/USPTO.spool.p/US10791017/runat.17022005.125807.22086/app.query.fasta_1.1358
-DB=EST -QFMT=faaap -SUFFIX=p2n.rst -MINMATCH=0.1 -LOOPT=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blomsum62 -TRANS=human40.cdd -LIST=45
-LOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pco -NORM=ext -HEASize=500 -MINLEN=0 -MAXLEN=200000000
-USER=US10791017@cgn.1.1.6628@runat.17022005.125807.22086 -NCPU=6 -ICPU=3
-NO MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -NARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

EST: *
1: gb_esc1:*
2: gb_esc2:*
3: gb_hic:*
4: gb_esc3:*
5: gb_esc4:*
6: gb_esc5:*
7: gb_esc6:*
8: gb_gss1:*
9: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1922	100.0	2073	CR593724	full-length
2	1922	100.0	2083	CR608666	full-length
3	1922	100.0	2103	CR619493	full-length
4	1922	100.0	2157	CR604952	full-length
5	1922	100.0	2314	CR608375	full-length
6	1907.5	99.2	2070	CR602561	full-length
7	1907.5	99.2	2093	CR594868	full-length
8	1907.5	99.2	2126	CR610888	full-length
9	1907.5	99.2	2159	CR625247	full-length

10	1888.5	98.3	2107	3	AK019460	AK019460	Mus muscu
11	1874	97.5	1379	3	AK014366	AK014366	Mus muscu
12	1871.5	97.4	2373	3	AK034755	AK034755	Mus muscu
13	1744.5	90.8	2236	3	CR604639	CR604639	full-length
14	1719.5	89.5	2269	3	AK049743	AK049743	Mus muscu
15	1675.5	87.2	4729	3	HSMB04802	AL833489	Homo sapi
16	1608.5	83.7	1787	3	CR618632	CR618632	full-length
17	1537.5	80.0	1895	5	EX408406	EX408406	EX408406
18	1537	80.0	1020	5	EX387336	EX387336	EX387336
19	1496.5	77.9	929	5	BQ881305	BQ881305	AGENCOURT
20	1495.5	77.8	1068	5	EX449773	EX449773	EX449773
21	1495	77.8	843	7	CN162640	CN162640	952042 MA
22	1482	77.1	897	6	BY713354	BY713354	BY713354
23	1470.5	76.5	1057	5	EX353312	EX353312	EX353312
24	1464	76.2	962	5	EX407870	EX407870	EX407870
25	1456.5	75.8	867	4	B1647600	B1647600	603278181
26	1450.5	75.5	940	5	BQ927930	BQ927930	AGENCOURT
27	1435	74.7	1013	5	EX465099	EX465099	EX465099
28	1435	74.7	1048	5	EX381313	EX381313	EX381313
29	1429	74.3	838	4	BG574437	BG574437	602596384
30	1428	74.3	1028	5	EX399069	EX399069	EX399069
31	1406	73.2	778	4	BG756215	BG756215	602713525
32	1403	73.0	1055	5	EX428311	EX428311	EX428311
33	1400.5	72.9	1026	5	BQ060808	BQ060808	AGENCOURT
34	1390	72.3	1022	5	EX407825	EX407825	EX407825
35	1386	72.1	792	5	EX368129	EX368129	EX368129
36	1377.5	71.7	837	4	BG751179	BG751179	60272679
37	1377	71.6	744	7	CR940579	CR940579	4063812 B
38	1373	71.4	925	5	EX333942	EX333942	EX333942
39	1372	71.4	933	5	EX387305	EX387305	EX387305
40	1368.5	71.2	2632	3	BC043904	BC043904	Xenopus 1
41	1368	71.2	901	2	BE562296	BE562296	601344870
42	1363	70.9	715	5	BQ042484	BQ042484	UI-M-EMO
43	1362	70.8	737	5	EX919060	EX919060	EX919060
44	1361	70.9	828	5	BQ671992	BQ671992	AGENCOURT
45	1347	70.1	787	2	BE741903	BE741903	601594075

ALIGNMENTS

RESULT 1
CR593724
LOCUS
DEFINITION
full-length cDNA clone CS0D1011Y005 of Placenta Cot 25-normalized
of Homo sapiens (human).
ACCESSION
CR593724
VERSION
CR593724.1 GI:50474531
KEYWORDS
HTC; CNSIT_CDNA.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens

REFERENCE
AUTHORS
Li, W.B., Gruber, C., Jessee, J. and Polyes, D.
TITLE
Full-length cDNA libraries and normalization
JOURNAL
Unpublished
REMARK
Contact: Feng Liang Email: fliang@life.techn.com URL:
http://fulllength.invitrogen.com/ invitrogen Corporation 1600
Paraday Avenue
2 (bases 1 to 2073)

REFERENCE
AUTHORS
TITLE
JOURNAL
Direct Submission
Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)

COMMENT
1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
end enriched, double-strand cDNA was digested with Not I and cloned
into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
was normalized. Library was constructed by Life Technologies, a
division of Invitrogen.

FEATURES
source
1..2073
/organism="Homo sapiens"

ORIGIN

/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0D1011Y005"
/issue_type="Placenta Cot 25-normalized"
/plasmid="pCMVSPORT_6"

Alignment Scores:

Pred. No.: 4,11e-109 Length: 2073
Score: 1922.00 Matches: 338
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
Gaps: 0

US-10-791-017a-2_copy_319_656 (1-338) x CR593724 (1-2073)

QY 1 G1YGLYARGLYGLYMETGLYSERIALAGLYGLUARGLYGLYPHEANLYSPROGLYGLY 20
DB 906 GGAGGACGGGTGGATGGGCGCTGGAGAGCGAGTGGCTTCATTAAGCCTGGTGA 965
QY 21 PROMETASPGIUGLYPROASPLEUASPLEUGLYPROBROVALASPROASPGIUSPSE 40
DB 966 CCATGATGATGAGGACCAATCTTGATCTAGGCCCACTGTAGATCCAGATGAAGACTCT 1025
QY 41 ASPASNSERIALLEYVALINGLYLEUASNSPSEVALTHLEUASPSPLEUALA 60
DB 1026 GACAAACGTCATTTATGTACAAAGATTAATGACGTGTGACTCTAGATGATCTGGCA 1085
QY 61 ASPSPHELYSGINCYSGIYVALIYLYMEASNSPSEVALTHLEUASPSPLEUALA 80
DB 1086 GACTTCTTTAAGCAGTGTGGGCTTTAAGTGAACAAAGAACTGGGCAACCATATC 1145
QY 81 H1E1EYTRLEUASPLYGLUTRGLYLYSPROLYSGLYASPA1ATHVALISERTYGLU 100
DB 1146 CACATCTACCTGGCAAGAAACAGAAAGCCCAAGAGCATGCGCAAGTCTCTATGAA 1205
QY 101 ASPPROBOTHRALVALVALAVALAVALGLUTRPHASPGIYLYSASPHEGLINGLYSER 120
DB 1206 GACCCACCCCACTGGCAAGGCTGGCGTGAATGATTTATGGGAAAGATTTTCAAGGGAGC 1265
QY 121 LYSELYSVYALISERLEUALAARGLYLYSPROBROMEASNSERMETARGLYGLYLEU 140
DB 1266 AATCTTAAGTCTCTCTTCTCTCGAAGAGCCTCAATGAACATTAAGCGGGGTGCTG 1325
QY 141 PROBROARGIUGLYARGLYMETPROBROBROLEUARGLYGLYPROGLYGLYPROGLY 160
DB 1326 CCAACCCGTCGAGGCAAGGAGCATCCACCACTCGGTGAGGTCCAGAGGCGCCAGGA 1385
QY 161 GLYPROGLYGLYPROMETGLYARGMETGLYGLYARGGLYGLYASPAARGLYGLYPHEPRO 180
DB 1386 GGTCTGGGGGAGCCCATGGGTGCGCATGGAGCGCGTGGAGAGATTAAGGAGGCTTCCCT 1445
QY 181 PROARGIYPROARGIYSEYERGLYASNSPSEYERGLYGLYAAVALGHN1AARG 200
DB 1446 CCAAGAGGACCCCGGGGTTCAGAGGAGAACCTCTGGAGAGAGAAAGTCCACACCGGA 1505
QY 201 ALAAGLYASPTIPGLYCYSPROASNSPGLYCYSGIYASNSPHEALATPARGTHR 220
DB 1506 GCTGGAGACTGGCAGTGTCCCAATCCGGGTGTGGAAACAGAACTTCGCTGGAGAAACA 1565
QY 221 GLUCYSAANGINCYLYVALAIPROLYSPROGLYGLYPHELEUPROPROPHOPHEPRO 240
DB 1566 GAGTGAACACAGTGAAGGCCCAAGCCTGAAGGCTTCTCCGCGCACCTTTCCGCGCC 1625
QY 241 PROGLYGLYASPAARGIYARGGLYGLYPROGLYGLYMETARGGLYGLYARGGLYGLYLEU 260
DB 1626 CCGGAGTGTGATCGTGGCAGAGGTGGCCCTGTGTGCAAGCGGAGAGAGGTGGCTC 1685
QY 261 METASPAARGIYGLYPROGLYGLYMETPHEARGGLYGLYARGGLYGLYASPAARGGLYGLY 280
DB 1686 ATGATCTGTGTGTCCCGGTGGATGTTCAGAGGTGGCCGTGTGGAGAGAGAGGTGGC 1745

QY 281 PHEARGGLYGLYARGGLYMETASPAARGIYGLYPHEGLYGLYGLYARGARGGLYGLYPRO 300
DB 1746 TTCGGTGTGGTGGCGGGGCAATGACAGAGTGGCTTTGGTGGAGAGAGAGAGTGGCCCT 1805
QY 301 G1YGLYPROPROGLYPROLEUWETGLUINMETGLYGLYARGARGGLYGLYARGGLYGLY 320
DB 1806 GGGGGGGCCCCCTGGACCTTTGATGGAACAGATGGGAGAGAAAGAGAGAGACCTGGAGGA 1865

QY 321 PROGLYLYMEASPLYGLYGLYH1SARGIINGLYARGARGASPARPROTYR 338
DB 1866 CCTGAAAAATGATTAAGCGAGCACGTCAGAGCGGAGAGATCGCCCTAC 1919

RESULT 2
CR608666
LOCUS
DEFINITION full-length cDNA clone CS0D004Y123 of B cells (Ramos cell line) of Homo sapiens (human).
ACCESSION CR608666
VERSION CR608666.1 GI:50489473
KEYWORDS HTC; CNSTL cDNA.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 2083)
AUTHORS Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished
REMARK Contact: Feng Liang Email: filiang@lifetech.com URL: http://fulllength.invitrogen.com/ Invitrogen Corporation 1600 Faraday Avenue
2 (bases 1 to 2083)
REFERENCE
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (20-JUN-2004) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail: seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies, a division of Invitrogen.
FEATURES
source location/Qualifiers
1..2083
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0D004Y123"
/issue_type="B cells (Ramos cell line)"
/plasmid="pCMVSPORT_6"

ORIGIN

Alignment Scores:

Pred. No.: 4,14e-109 Length: 2083
Score: 1922.00 Matches: 338
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
Gaps: 0

US-10-791-017a-2_copy_319_656 (1-338) x CR608666 (1-2083)

QY 1 G1YGLYARGLYGLYMETGLYSERIALAGLYGLUARGLYGLYPHEANLYSPROGLYGLY 20
DB 906 GGAGGACGGGTGGATGGGCGCTGGAGAGCGAGTGGCTTCATTAAGCCTGGTGA 965
QY 21 PROMETASPGIUGLYPROASPLEUASPLEUGLYPROBROVALASPROASPGIUSPSE 40
DB 966 CCATGATGATGAGGACCAATCTTGATCTAGGCCCACTGTAGATCCAGATGAAGACTCT 1025
QY 41 ASPASNSERIALLEYVALINGLYLEUASNSPSEVALTHLEUASPSPLEUALA 60
DB 1026 GACAAACGTCATTTATGTACAAAGATTAATGACAGTGTGACTCTAGATGATCTGGCA 1085

OY		61	AppPhePhyLysGlcysGlyValIallylMetCantLysArgThSGLVLPrometile	80
Db		1086	GACTCTTTAGCAAGTGTGGGTTTGTAATGAACAGAACTGGGCACCATTATTC	1145
OY		81	HisieryLeuapLysGluThrGlyLysProLySGlyAspAlaThrValSerTyrglu	100
Db		1146	CACATCTACCTGACAGAGAAACAGAAAAGCCCAAAGGGAGATGCCAAGTGTCTATGA	1205
OY		101	AAPPROPTHAlAlalyAlaValGluTPHieapRglyLysaspPheGlnGlyser	120
Db		1206	GACCAACCACCGCACAGAGCTGCGCTGGATGATGTATGGGAAAGATTTCAGAGGAGC	1265
OY		121	LysLeuLysValSerLeuAlaArgLysLysProPrometAanSemetArgGlyVgLYeu	140
Db		1286	AAACTTTAAAGTCTCCCTTGCTCGAANAAGCTCCAAATGAACAGATATCGGGGTGTCTG	1325
OY		141	ProProArgGluGlyArgGlyLmetProProProLeuArgGlyVgLYProGlyVgLYProGly	160
Db		1326	CCACCCCCGAGAGGAGGAGCATGCCCACCACTCCGAGAGGTCCAAGAGGCCACAGA	1385
OY		161	GlyProGlyVgLYPrometGlyArgmetGlyVgLYArgGlyVgLYasparGlyVgLYPhePro	180
Db		1386	GATCTCGGGAGACCCATGGGTGTGCATGGAGGCGCTGAGAGGATAGAGAGGCTTCTCT	1445
OY		181	ProARsgLYPrArgGlySerArgLysAnPProSerGlyVgLYVAsnValGlnHisArg	200
Db		1446	CCMAAGAACCCCCGGGGTTCCTCCAGAGGAACCTCTCGAGAGAGAAAGCTCCAGCACGA	1505
OY		201	AlagIyaPTripInCyseProanProGlyCySGlyAenGlnasnPheAlaTParGthr	220
Db		1506	GCTGGAGACTGGCAGTGTCCCAATCCGGGTTGTGAAACAGAACTTGCGCTGAGAAACA	1565
OY		221	GluCySaenGlnCyLysAlaProLysProGluGlyPHeleuProProProPheProPro	240
Db		1566	GAGTCAACCAAGTGTAAAGCCCCCAAGCCTGAAGGCTTCTCCGCCACCTTTCCGGCC	1625
OY		241	ProGlyVgLYasParGlyVArgGlyVgLYProGlyVgLYmeCtArGlyVgLYVArgGlyVgLYeu	260
Db		1626	CCGGGTGTGATCTGTGGCAGAGGTGGCTTGTTGGCATTCGGGGAGAAAGAGTGGCTTC	1685
OY		261	MetaPaArgGlyVgLYProGlyVgLYmetPheaArgGlyVgLYArgGlyVgLYasParGlyVgLY	280
Db		1686	ATGATCTGTGGGTGTCCGGTGGATGTTCAGAGGTGGCCGGTGGAGAGACAGAGTGGC	1745
OY		281	PheaArgGlyVgLYArgGlyLmetAspaArgGlyVgLYPheGlyVgLYVgLYVArgGlyVgLYPro	300
Db		1746	TTCCTGTGTGGCCGGGAGCATGGACAGAGTGGCTTTGGTGGAGGAACGAGGTGGCCCT	1805
OY		301	GlyGlyProProGlyVgLYProLewmetGlnGlnmetGlyVgLYVArgArgGlyVgLYVArgGlyVgLY	320
Db		1806	GGGGGGCCCCCTCGACCTTTGATGTGAACAGATGGAGAGAAAGAGAGACGTGGAGGA	1865
OY		321	ProGlyLysMeLaspLysGlyVgLYHisArgGlnIuaArgArGaSpaArgProTyR	338
Db		1866	CCTGGAATAATGATTAAGGCGAGCACCGTCAAGAGCCGACAGATCGGCCCTTAC	1919
RESULT 3				
LOCUS	CR619493	2103 bp	mRNA	linear HTC 21-JUN-2004
DEFINITION	Full-length cDNA clone CSOD1076Y021 of Placenta Cot 25-normalized			
ACCESSION	CR619493			
VERSION	CR619493.1	GI:50500300		
KEYWORDS	HTC; CNSLT_cDNA.			
SOURCE	Homo sapiens (human)			
ORGANISM	Homo sapiens			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
TITLE	1.(bases 1 to 2103)			
JOURNAL	L4.W.B., Gruber,C., Jeejee,J. and Polayes,D.			
	Full-length cDNA libraries and normalization			
	Unpublished			

REMARK	COMMENT	FEATURES
<p>CONTACT : FENG LIANG EMAIL : fliang@lifeitech.com URL : http://fulllength.invitrogen.com/ INVITROGEN CORPORATION 1600 PARADISE AVENUE 2 (BASES 1 TO 2103)</p>	<p>Submitted (20-JUN-2004) Genoscope - Centre National de Séquençage : BP 191 91006 Evry cedex - FRANCE (E-mail : secref@genoscope.cns.fr)</p> <p>- Web : www.genoscope.cns.fr</p> <p>1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies, a division of Invitrogen.</p>	<p>Location/Qualifiers</p> <p>1..2103</p> <p>/organism="Homo sapiens"</p> <p>/mol_type="mRNA"</p> <p>/db_xref="taxon:9606"</p> <p>/clone="CS0D10761021"</p> <p>/tissue_type="Placenta Cot 25-normalized"</p> <p>/plasmid="pCMVSPORT_6"</p>
<p>ALIGNMENT SCORES:</p> <p>Pred. No. : 4,18e-109</p> <p>Score: 1922.00</p> <p>Percent Similarity: 100.00%</p> <p>Best Local Similarity: 100.00%</p> <p>Query Match: 100.00%</p> <p>Ds: 3</p> <p>Gaps: 0</p>	<p>Length: 2103</p> <p>Matches: 338</p> <p>Conservative: 0</p> <p>Mismatches: 0</p> <p>Indels: 0</p>	<p>US-10-791-017A-2_COPY_319_656 (1-338) x CR619493 (1-2103)</p>
<p>1 GLVGIYARVGIYGLYMETGLYSERIALGLVGLIUAIRVGIYGLYPHEASNTYSPROGLYGLY 20</p> <p>943 GGAGGACCCGGTGGATGATGGCGAGCGCTGGAAGAGGAGGAGGTGGCTTCATTAAGACCTGGTGG 1002</p> <p>21 PROMETASPGIUGLYPROASPLEUASPLEUGLYPROPROVALASPROMASPGIUNASPER 40</p> <p>1003 CCCATGGATGAAGGACCAAGCATCTTGATGTAGGCCCACTGTATGATCCGATGAAGACCTCT 1062</p> <p>41 ASPANSERIALLETYYVALGLIGLYLEUASNASPSEVALTHIRLEUASPLEUAA 60</p> <p>1063 GACAAACAGTGCATTTATGTACAAGGATTAATGACAGTGTGACTCTGATGATCTGGCA 1122</p> <p>61 ASPHEPHELYSGINCYSGLYVALVALLYMETASNTLYSAAGTHIRGLYGLINPROMETILE 80</p> <p>1123 GACTTCCTTTAAGCAGGTGTGGGGTGTTAAGATGAAACAAGAACTGGGCAACCATGATC 1182</p> <p>81 HISILETYRLEUASPLYGLUITHGLYLYSPROBYSGLYASPAIATNRVALSERTYGLU 100</p> <p>1183 CACATCTAACCTCGGACAAAGAAACAGAAAGCCCAAAGGCATGCCACAGTCTCTATGAA 1242</p> <p>101 ASPPROPTHIRALALYALALALVALGLUTRIPHEASPGIYLYSASPHEGLINGLYSER 120</p> <p>1243 GACCCACCACTGGCAAGGCGTGGCGGTGAATGGTTTGAATGGAAAGATTTTCAAGGAGGC 1302</p> <p>121 LYSLEULYVALSERLEUAAIRGLYLYSPROBOMETANSEMETATYGLYGLYLEU 140</p> <p>1303 AAACCTTAAGCTCTCCCTGTCTCGGAAGAAGCCTCAATGAACAGTATAGCCGGGGTGCTTG 1362</p> <p>141 PROPROAHRVGIUGLYATVSGIYMETSPROPROBOMELUARGIYGLYPROGLYGLYPROGLY 160</p> <p>1363 CCACCCCGTGAAGGCAAGGAGCATCCACACCACTCCGTGAGAGTCCAGAGGCCCCAGGA 1422</p> <p>161 GLYPROGLYGLYPROMETGLYARGMETGLYGLYARGIYGLYASPARGLYGLYPHEPRO 180</p> <p>1423 GGTCTGGGGGACCCATAGGATCGCATGGAGGCGCGTGGAGGAGATGAGGAGGCTTCCCT 1482</p> <p>181 PROARGLYPROARGLYSERARGLYASNPROSERGLYGLYGLYASNVALGLNHIAIR 200</p> <p>1483 CCAAGAGGACCCCGGGGTTCCTCGAGGGAACCCCTCTGTGAGAGGAAACGTCTCACACCGA 1542</p>	<p>1 GLVGIYARVGIYGLYMETGLYSERIALGLVGLIUAIRVGIYGLYPHEASNTYSPROGLYGLY 20</p> <p>943 GGAGGACCCGGTGGATGATGGCGAGCGCTGGAAGAGGAGGAGGTGGCTTCATTAAGACCTGGTGG 1002</p> <p>21 PROMETASPGIUGLYPROASPLEUASPLEUGLYPROPROVALASPROMASPGIUNASPER 40</p> <p>1003 CCCATGGATGAAGGACCAAGCATCTTGATGTAGGCCCACTGTATGATCCGATGAAGACCTCT 1062</p> <p>41 ASPANSERIALLETYYVALGLIGLYLEUASNASPSEVALTHIRLEUASPLEUAA 60</p> <p>1063 GACAAACAGTGCATTTATGTACAAGGATTAATGACAGTGTGACTCTGATGATCTGGCA 1122</p> <p>61 ASPHEPHELYSGINCYSGLYVALVALLYMETASNTLYSAAGTHIRGLYGLINPROMETILE 80</p> <p>1123 GACTTCCTTTAAGCAGGTGTGGGGTGTTAAGATGAAACAAGAACTGGGCAACCATGATC 1182</p> <p>81 HISILETYRLEUASPLYGLUITHGLYLYSPROBYSGLYASPAIATNRVALSERTYGLU 100</p> <p>1183 CACATCTAACCTCGGACAAAGAAACAGAAAGCCCAAAGGCATGCCACAGTCTCTATGAA 1242</p> <p>101 ASPPROPTHIRALALYALALALVALGLUTRIPHEASPGIYLYSASPHEGLINGLYSER 120</p> <p>1243 GACCCACCACTGGCAAGGCGTGGCGGTGAATGGTTTGAATGGAAAGATTTTCAAGGAGGC 1302</p> <p>121 LYSLEULYVALSERLEUAAIRGLYLYSPROBOMETANSEMETATYGLYGLYLEU 140</p> <p>1303 AAACCTTAAGCTCTCCCTGTCTCGGAAGAAGCCTCAATGAACAGTATAGCCGGGGTGCTTG 1362</p> <p>141 PROPROAHRVGIUGLYATVSGIYMETSPROPROBOMELUARGIYGLYPROGLYGLYPROGLY 160</p> <p>1363 CCACCCCGTGAAGGCAAGGAGCATCCACACCACTCCGTGAGAGTCCAGAGGCCCCAGGA 1422</p> <p>161 GLYPROGLYGLYPROMETGLYARGMETGLYGLYARGIYGLYASPARGLYGLYPHEPRO 180</p> <p>1423 GGTCTGGGGGACCCATAGGATCGCATGGAGGCGCGTGGAGGAGATGAGGAGGCTTCCCT 1482</p> <p>181 PROARGLYPROARGLYSERARGLYASNPROSERGLYGLYGLYASNVALGLNHIAIR 200</p> <p>1483 CCAAGAGGACCCCGGGGTTCCTCGAGGGAACCCCTCTGTGAGAGGAAACGTCTCACACCGA 1542</p>	<p>US-10-791-017A-2_COPY_319_656 (1-338) x CR619493 (1-2103)</p>

Qy	201	AlaGlyAspTTPGInCysProAsnProGlyCysGlyAsnGluAsnPheAlaTyrArgThr	220
Db	1543	GCTGAGACTGCGACAGTGTCCCAATCCGGATTGTGTGAAACCAAGACTTCGCTGGAGAAACA	1602
Qy	221	GluCysAsnGlnCysIlyAspAlaProIlyAspProGlnIlyPheLeuProProPhePro	240
Db	1603	GAGTGCACCAAGTGTAAAGGCCCCAAAGCCTTGAAGGCTTCCTCCGCGACCTTTCGGCCCC	1662
Qy	241	ProGlnIlyIlyAspAspArgIlyIlyArgIlyIlyProGlnIlyIlyMetArgIlyIlyIlyArgIlyIlyLeu	260
Db	1663	CCGGGTGTGTATCGTGGCGAAGGTGGCCCTGTGTGGCAAGCGAGAGAAAGGTGGCTC	1722
Qy	261	MetAspArgIlyIlyIlyProGlnIlyIlyMetPheArgIlyIlyIlyArgIlyIlyIlyAspArgIlyIlyIly	280
Db	1723	ATGATTCGTGTGTGTGTCCTCCGTGTGAATGTTCACAGGTGGCCCTGTGTGAAGACAGAGGTGGC	1782
Qy	281	PheArgIlyIlyIlyArgIlyIlyMetAspArgIlyIlyIlyPheGlnIlyIlyIlyArgIlyIlyIlyPro	300
Db	1783	TTCCGTGTGTGGCCGGGGCATGTGAACCAAGTGTGCTTGTGTGAAGGAACGAGTGTGGCCCT	1842
Qy	301	GlnIlyIlyProProProGlnIlyProLeuMetGlnGlnMetGlnIlyIlyIlyArgIlyIlyIlyArgIlyIlyIly	320
Db	1843	GGGGGGCCCCCTGGACCTTTGATGTGAACAGATGGAGAGAAAGAGGAGGACGTGGAGGA	1902
Qy	321	ProGlnIlyIlyMetAspIlyAspGlnIlyIlyHisArgGlnIlyIlyArgIlyIlyIlyAspArgProTyr	338
Db	1903	CCTGGAATAATGTGATTAAGCGGAGACCACTGTACAGAGCCGACAGATCGGCTTAC	1956

RESULT 4					
CR604952					
LOCUS	CR604952	2157 bp	mRNA	linear	HTC 21-JUL-2004
DEFINITION	Full-length cDNA clone CS0D1010Y112 of Placenta Cot 25-normalized of Homo sapiens (human).				

ACCESSION	CR604952
VERSION	CR604952.1
KEYWORDS	GI:50485759
SOURCE	HTC; CNSLT_CDNA
ORGANISM	Homo sapiens (human)
	Homo sapiens

REFERENCE	AUTHORS	TITLE	JOURNAL	REMARK
1 (bases 1 to 2157)	Ll, W.B., Gruber, C., Jesse, J. and Polayes, D.	Full-length cDNA libraries and normalization	Unpublished	Contact : Feng Liang Email : fliang@lifecycle.com URL : http://www.lifecycle.com

REFERENCE 2 (bases 1 to 2157)

TITLE Direct Submission
JOURNAL Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage :

COMMENT
1st strand cDNA was primed with a NotI-ligo (dT)₁₈ primer. Five primers
end enriched, double-strand cDNA was digested with Not I and cloned
into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
was normalized. Library was constructed by Life Technologies, a
division of Invitrogen.

FEATURES	Location/Qualifiers
source	1. .2157

```

/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CSODI010Y112"
/tissue="placenta"
/plasmid="pCMVSPORT_6"

```

Alignment Scores:	4.3e-109	Length:	2157
Pred. No.:		Matches:	338
Score:	1922.00	Conservative:	0
Percent Similarity:	100.00%		

Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	3	Gaps:	0

US-10-791-017A-2_COPY_319_656 (1-338) x CR604952 (1-2157)

QY	1	GIYGLIYATRGILYGLMETGLYSERAIGLIYGLIUAHGRGLYGLIYPHEANLNPSPROGLIYGLY	20
Db	995	GGAGAGACGGGTGGAAATGGGCGAGCGCTGGAGAGCGAGAGGTGGCTTCAATTAAGCCCTGGTGGGA	1054
QY	21	PROMETASPGIUGIYPIROASPLEUASPLEUGIYPROBOVALIASPPOASPGIUAASPER	40
Db	1055	CCCATGTGATGAGAACCAAGACTCTTGATCTTAGGCCCACTGTGAGATCCAGATGAGAACTCT	1114
QY	41	ASPAASERAIILIEIYVALINGIYLEUASNASPSEVALIIRLEUASPAASPLEUA	60
Db	1115	GACAAACAGTCGCAATTTATGTACAGAGATTAAATGACAGTGTGACTTAGATGATGACTGGCA	1174
QY	61	ASPHPELHLYSGINCYGGLIYVALIYALVLYMEKLESLNYSARGTHIRGILYGLINPROMETILE	80
Db	1175	GACTTCTTTAAGCAGGTGTGGGGTGTTAAGATGAAACAAGAACTGGGCAACCCATATATC	1234
QY	81	HISILEIYRLIENUAPLYSGIUTHIRGLIYLSYPROBLYSGIYASPAIATHIRVALISERTYGLU	100
Db	1235	CACATCTTACCTGGACAAAGAAACAGAAAGCCCAAGGCCATGTCACATGTCCTTAAGAA	1294
QY	101	ASPPROTHIRAIALYLSAIAIAVALIGIUTHIRPHEASPGIYLYSAEPHENGILYSEY	120
Db	1295	GACCAACCCCACTGGCAAGGCTGCGGTGGAAATGTTTGATGGGAAAGATTTTCAAGGGAGNC	1354

QY	121	LysLeuLysValIserLeuAlaArgLysLysProPheMetLeuSerMetArgLysLysLeu	140
Db	1355	AAACCTAAAGCTCCCTTGTCTGGAGAAAGCTCCAAATGAACACTATAGCGGGAGGTGTG	1414
QY	141	ProProArgLysLysLysArgLysMetProProPheLeuArgLysLysProLysLysProLys	160
Db	1415	CCACCCCTGAGGGCAGAGGCATGCCACACCACTCCGTGGAGGTCCAGAGAGGCCACAGA	1474
QY	161	GlyProGlyLysProMetGlyLysMetGlyLysLysArgLysLysAspArgLysLysPhePro	180
Db	1475	GGTCTCGGGGAGCCCATGGGTGCGATGGAGGCCGTGGAGAGATAGAGGAGGCTTCCCT	1534
QY	181	ProArgGlyProArgLysSerArgGlyAsnProSerGlyLysLysAsnValGlnHisArg	200
Db	1355	CCAGAGGACCCCGGGGTTCCCGAGGAAACCCCTCTGGAGAGAGAAAGCTCCACACCGA	1594
QY	201	AlaGlyAspTrpGlnCysProAsnProGlyCysGlyAsnGlnAsnPheAlaTrpArgThr	220
Db	1595	GCTGGAGACTGGCAGGTGTCCCAATCCGGGGTGTGGAAACCGAACTTCGCTGGAGAAC	1654
QY	221	GlyCysAsnGlnCysLysValAlaProLysAspProLysLysPheLeuProProProPheProPro	240
Db	1655	GAGTGCACACAGGTGTAGAGGCCCCAAAGCTTAAGGCTTCTCCGCGCACCCCTTTCGGCC	1714
QY	241	ProGlyLysAspArgLysArgLysLysProLysLysMetArgLysLysArgLysLysLeu	260
Db	1715	CCGGGTGTGATCGGTGGCAGAGGTGGCCCTGTGTGGCATGGGGAGGAAGAGGTGGCCTC	1774
QY	261	MetAspArgLysLysLysProGlyLysMetPheArgLysLysLysArgLysLysAspArgLysLys	280
Db	1775	ATGATATCGTGTGTGCTCCGGTGAATGTTCACAGGTGGCCGTGTGGAGACACAGGTGGC	1834
QY	281	PheArgLysLysArgLysMetAspArgLysLysPheGlyLysLysArgLysLysLysPro	300
Db	1835	TTCCCTGTGGCCGGGGCATGGAACGAGGTGGCTTTGTGGAGGAAGCAGAGTGGCCCT	1894
QY	301	GlyLysProProGlyProLeuMetGlnLysMetGlyLysArgLysLysLysLysLysLysLys	320
Db	1895	GGGGGGCCCCCTGGGACTTTGATGGACAGATGGAGAGAAAGAGGAGACGTGGAGGA	1954
QY	321	ProGlyLysMetAspLysGlyLysHisArgLysLysArgLysLysArgLysProLys	338
Db	1955	CTTGAAATAATGATTAAGCGAGACACCTGTACAGAGCGCAGAGATCGGCCCTTAC	2008

LOCUS	CR608375	2314 bp	mRNA	linear	HTC 21-JUL-2004
DEFINITION	full-length cDNA clone CSODI039Y24 of Placenta Cot 25-normalized of Homo sapiens (human).				
ACCESSION	CR608375				
VERSION	CR608375.1 GI:50489182				
KEYWORDS	HTC; cNSLT cDNA.				
SOURCE	Homo sapiens (human)				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
AUTHORS	1 (bases 1 to 2314)				
TITLE	Li, W.B., Gruber, C.J., Jessup, J. and Polayes, P.				
JOURNAL	Full-length cDNA libraries and normalization				
REMARK	Unpublished				
REFERENCE	Contact : Feng Liang Email : fliang@life.technet.com URL : http://fulllength.invitrogen.com/ Invitrogen Corporation 1600				
AUTHORS	Paradey Avenue				
TITLE	2 (bases 1 to 2314)				
JOURNAL	Genoscope.				
REMARK	Direct Submission				
COMMENT	Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage BP 191 91006 EVRY cedex - FRANCE (E-mail : sequef@genoscope.cns.fr				
FEATURES	1st strand cDNA was primed with a NotI-oligo (dT) primer. Five primers end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by life technologies, a division of invitrogen.				
FEATURES	Location/Qualifiers				
SOURCE	1..2314				
ORIGIN	/organism="Homo sapiens"				
	/mol_type="mRNA"				
	/db_xref="taxon:9606"				
	/clone="CSODI039Y24"				
	/tissue_type="Placenta Cot 25-normalized"				
	/plasmid="pCMVSPORT_6"				
ALIGNMENT SCORES:					
Pred. No.:	4.63e-109	Length:	2314		
Score:	1922.00	Matches:	338		
Percent Similarity:	100.00%	Conservative:	0		
Best local Similarity:	100.00%	Mismatches:	0		
Query Match:	100.00%	Indels:	0		
DB:	3	Gaps:	0		
US-10-791-017A-2_COPY_319_656 (1-338) x CR608375 (1-2314)					
QY	1	GLYGLYATAGGGLYGLYMERGLYSERIALGLYLUATRGGLYGLYPLEAANTLYSPROGLYGLY	20		
DB	995	GGAGGACCGGGTGGAAATGGCGCGCTGGAGACGAGGTGGCTTCATTAAGCCCTGGTGA	1054		
QY	21	PROMETASPGIUGLYPRCASPPLAUSPPLUGLYPROPVALAASPPTCASPGLAASPSE	40		
DB	1055	CCCATGGATGAAGGACGACGACTTGATGTAGGCCCACTGTAGATCCAGATGAAGACTCT	1114		
QY	41	ASPANSERIALIETRYVALGLINGLYLEUANAASPSEVALTHYLEAASPPLAULA	60		
DB	1115	GACAAACGTGCATTTATGTACAAAGATTAATGACAGTGCATCTAGATGATCTGGCA	1174		
QY	61	ASPHEPEHYEGLYCYSGLYVALVALYMERANLYSAGTTHRGYGLINPROMETILE	80		
DB	1175	GACTTCTTTAAGCAGGTGTGGGGTGTTAGTAGTAACAAGAACTGGGCAACCATATTC	1234		
QY	81	HISILETRYLEAUSPGLYGLYTHRGYLYSPFOLYSGLYAEPALATINVALSERTRYGLU	100		
DB	1235	CACATCTACCTGGACAAAGAAACAGGAAAGCCAAAGCGATGCGACAGTGTCTTAGAA	1294		
QY	101	ASPPTPOTTHRALYVALALALAVAGLUITRPHASPGYLYVSPAPHEGLINGLYSER	120		

Accession	Definition	Source	Organism	Reference Authors Title Journal	Comment
Db	1295 GACCCACCACTGGCCCAAGGCTGCCGTGAATGGTTGATGGGAAAGATTTCACGAGGAGC				
QY	121 LysleuLysVal1SerLeuAlaArguYelYsProPomEtanSerMetArgGlyGlyLeu				
Db	1355 AAACCTTAAGTCTCCCTTGCTCTGGAGAAGGCTCCAAATGAACAGATCCGGGGTGGTCTG				
QY	141 ProPArGtGluGlyArgGlyMetProProProLeuArGlyGlyProGlyGlyProGly				
Db	1415 CCACCCCGTGAAGGGGAGAGGCGATGCCACCACTCCGTGAGAGGTCCAGAGAGGCCAAGA				
QY	161 GlyProGlyGlyProMetGlyArgMetGlyGlyArgGlyGlyAspArgGlyGlyPhePro				
Db	1475 GGTCTCGGGGAGCCCATGGGTCTGCATGGAGGCGCGTGAAGAGATAGAGGAGGCTTCCCT				
QY	181 ProArGtGlyProArGtGlySerArGtGlyAsnProSerGlyGlyGlyAsnValGlnHsArg				
Db	1555 CCAGAGAGACCCCGGGGATTCCTCCAGAGAACCCCTCTGGAGAGAGAAAGTCCAGACCGCA				
QY	201 AlaGlyAspTrpGlnCysProAsnProGlyCysGlyAsnGlnAsnPheAlaTrpArgThr				
Db	1595 GCTGGAGACTGGCAGGTGCCCAATCCGGGTTGTGGAAACCAAGACTTCGCTGGAGAACCA				
QY	221 GluCysAsnGlnCysGlyValAlaProLysProGluGlyPheLeuProProProPheProPro				
Db	1655 GAGTCAACCAAGTGAAGGCCCAAGGCTTAAGGCTTCTCCGCCACCTTTCCGCC				
QY	241 ProGlyGlyAspArgGlyArgGlyGlyProGlyGlyMetArgGlyGlyArgGlyGlyLeu				
Db	1715 CCGGTGGTATTCGTGGAGAGAGTGGCCCTGTGTGGCATGCCGGAGAGAAAGTGGCTT				
QY	261 MetAspArgGlyGlyProGlyGlyMetPheArgGlyGlyArgGlyGlyAspArgGlyGly				
Db	1775 ATGATCTGTGGTGTCCCGGTGAATGTTCAAGAGTGGCCGTGTGTGAACAGAGTGGC				
QY	281 PheArGtGlyGlyArgGlyMetAspArgGlyGlyPheGlyGlyGlyArgGlyGlyPro				
Db	1835 TTCCGTGTGGCCCGGGGATGATGACCGAGTGGCTTTGTGTGAAGAAACGAGTGGCCCT				
QY	301 GlyGlyProProGlyProLeuMetGluGlnMetGlyGlyArgArgGlyGlyArgGlyGly				
Db	1895 GGGGGGCCCTCGGACCTTTGATGAACAGATGGAGAGAGAAAGAGAGCAGTGAAGA				
QY	321 ProGlyLysMetAspLysGlyGlnHisArgGlnLysArgAspArgProTyr				
Db	1955 CCTGAAATAATGATTAAGGCGACACCGTTCAGAGCGCAGAGATCGGCTTAC				
RESULT 6	CR602561	2070 bp	mRNA	linear	HTC 21 -JUL-2004
LOCUS	CR602561	full-length cDNA clone	CSDD1068Y23	of Placenta cot 25-normalized	
ACCESSION	CR602561	of Homo sapiens (human).			
VERSION	CR602561.1	GI:50483368			
KEYWORDS	HTC, cNSLT cDNA.				
SOURCE	Homo sapiens (human)				
ORGANISM	Homo sapiens				
REFERENCE	Bukaryota, Metazoa, Chordata, Craniata, Vertebrata, Euteleostomi, Mammalia, Eutheria, Primates, Catarrhini, Hominidae, Homo.				
AUTHORS	1 (bases 1 to 2070)				
TITLE	Li, W.B., Gruber, C., Jessee, J. and Polayes, D.				
JOURNAL	Full-length cDNA libraries and normalization				
REMARK	Unpublished				
	Contact: Feng Liang Email: fliang@lifetech.com URL: http://fulllength.invitrogen.com/Invitroden Corporation 1600				
	Paradise Avenue				
	2 (bases 1 to 2070)				
	Genoscope.				
	Direct Submission				
	Submitted (20-JUN-2004) Genoscope - Centre National de Sequenage :				
	BP 191 9106 EVRY cedex - FRANCE (E-mail: seque@genoscope.cns.fr				
	- Web : www.genoscope.cns.fr)				
	1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime				
	end enriched, double-strand cDNA was digested with Not I and cloned				

into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies, a division of Invitrogen.

FEATURES

Source

Location/Qualifiers
1. .2070
/organism="Homo sapiens"
/mol_type="rRNA"
/db_xref="taxon:9606"
/clone="CS01064YE15"
/issue_type="Placenta Cot 25-normalized"
/plasmid="pCMVSPORT_6"

ORIGIN

Alignment Scores:

Score: 3.24e-108 Length: 2070
Percent Similarity: 1907.50 Matches: 337
Best Local Similarity: 99.70% Conservative: 0
Query Match: 99.25% Mismatches: 0
DB: 3 Indels: 1
Gaps: 1

US-10-791-017a-2_COPY_319_656 (1-338) x CR602561 (1-2070)

1 G1G1YARGLYGLYMETGLYSERIALAG1G1UARGLYGLYPHEANLYSPROGLYGLY 20
943 GGAGAGCGCGGTGGAATGGC---GCTGGAGAGCGAGGTGGCTTCATAGCCTGGTGA 999
21 PromeAapglu1g1yProAapleuAapleu1g1yProProValAapProAapSer 40
1000 CCCATGATGAGAGGACAGATCTTGATCTAGTCCACCTTAGATCCAGATGAAGACTCT 1059
41 AapAapSer1a1e1yrrval1g1ng1y1euanAapSerVal1h1r1euanAapleuA1a 60
1060 GACACACGTGCAATTATATGACAGATTAATATACAGTGTGACTTGAATGATCTGGCA 1119
61 AapAapAapleu1g1y1euanAapleuA1a1e1yrrval1g1ng1y1euanAapSer 80
1120 GACTTCTTAAAGCAGTGTGGGTGTTTAAATGATGAACAGAGAACTGGGCAACCATGATC 1179
81 H1e1e1yrrval1g1ng1y1euanAapSerVal1h1r1euanAapleuA1a1e1yrrval 100
1180 CACATCTACCTGCGCAAGAAACAGAAAGCCCAAGCCGATGCGACAGTCTCTATGAA 1239
101 AapAapProAapleuA1a1e1yrrval1g1ng1y1euanAapSerVal1h1r1euanAap 120
1240 GACCCACCACTGCGCAAGAGCTGCGGTGATGATGAGAAATTTTCAAGGAGC 1299
121 Lys1euanAapleuA1a1e1yrrval1g1ng1y1euanAapSerVal1h1r1euanAap 140
1300 AAATCTTAAGTCTCCCTTCTCGGAGAGAGCTTCATGAAACATATGCGCGGTGCTG 1359
141 ProAapAapleuA1a1e1yrrval1g1ng1y1euanAapSerVal1h1r1euanAap 160
1360 CCACCCCGTGAAGGAGGAGGATCCACACACCTCGTGAAGGAGGAGGAGGAGGAGGAG 1419
161 G1yProAapleuA1a1e1yrrval1g1ng1y1euanAapSerVal1h1r1euanAap 180
1420 GGTCTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1479
181 ProAapAapleuA1a1e1yrrval1g1ng1y1euanAapSerVal1h1r1euanAap 200
1480 CCAAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1539
201 A1a1e1yrrval1g1ng1y1euanAapSerVal1h1r1euanAapSerVal1h1r1euan 220
1540 GCTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1599
221 G1yProAapleuA1a1e1yrrval1g1ng1y1euanAapSerVal1h1r1euanAap 240
1600 GAGTGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1659
241 ProAapAapleuA1a1e1yrrval1g1ng1y1euanAapSerVal1h1r1euanAap 260

1660 CCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1719
261 MetAapAapleuA1a1e1yrrval1g1ng1y1euanAapSerVal1h1r1euanAap 280
1720 ATGAGTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1779
281 PheaAapAapleuA1a1e1yrrval1g1ng1y1euanAapSerVal1h1r1euanAap 300
1780 TTCGGTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1839
301 G1yProAapleuA1a1e1yrrval1g1ng1y1euanAapSerVal1h1r1euanAap 320
1840 GCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1899
321 ProAapAapleuA1a1e1yrrval1g1ng1y1euanAapSerVal1h1r1euanAap 338
1900 CCTGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1953
RESULT 7
LOCUS CR594868
DEFINITION full-length cDNA clone CS01064YE15 of Placenta Cot 25-normalized of Homo sapiens (human).
ACCESSION CR594868
VERSION CR594868.1 GI:50475675
KEYWORDS HTC; CNSLT cDNA.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 2093)
AUTHORS Li, W.B., Gruber, C., Jessee, J., and Polayes, D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished
CONTACT : Feng Liang Email : fliang@life.techn.com URL : http://fulllength.invitrogen.com/ Invitrogen Corporation 1600 Faraday Avenue
2 (bases 1 to 2093)
REFERENCE Genoscope.
AUTHORS Direct Submission
TITLE Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)
JOURNAL - Web : www.genoscope.cns.fr)
COMMENT 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies, a division of Invitrogen.
FEATURES
Source
Location/Qualifiers
1. .2093
/organism="Homo sapiens"
/mol_type="rRNA"
/db_xref="taxon:9606"
/clone="CS01064YE15"
/issue_type="Placenta Cot 25-normalized"
/plasmid="pCMVSPORT_6"

ORIGIN

Alignment Scores:

Score: 3.28e-108 Length: 2093
Percent Similarity: 1907.50 Matches: 337
Best Local Similarity: 99.70% Conservative: 0
Query Match: 99.25% Mismatches: 0
DB: 3 Indels: 1
Gaps: 1

US-10-791-017a-2_COPY_319_656 (1-338) x CR594868 (1-2093)

1 G1G1YARGLYGLYMETGLYSERIALAG1G1UARGLYGLYPHEANLYSPROGLYGLY 20
943 GGAGAGCGCGGTGGAATGGC---GCTGGAGAGCGAGGTGGCTTCATAGCCTGGTGA 999
21 PromeAapglu1g1yProAapleuAapleu1g1yProProValAapProAapSer 40

Db	1000	CCCATGATGTAAGGACACGACTCTTGATCTTAAGGCCACCTGTAGATCCAGATGAAGACTCT	1059
Qy	41	Aapbansera1a11eTyVa1GInG1yLeuembapserVal1ThrLeuAapbApleuA1a	60
Db	1060	GACAAACATGTCAATTATATGACMAAGATTTAAATGACAGTGAAGCTTAAGAATCTGCA	1119
Qy	61	AapPhepelysGInCyG1yVa1Val1ySmecAenlybAsGThnG1yGInProMec11e	80
Db	1120	GACTCTTTAAGCAGGTGTGGGGTTGTTAAATGAAACAGAGAACTGGGCACCCATGATC	1179
Qy	81	Hi11eTy1rLeuAapblyG1uThrG1yLybProlybG1yAapA1aThrVal1SerTyG1u	100
Db	1180	CACATCTTACCTGGACMAAGAAACAGAAAGCCCAAGGCCATGTCACAGTCTCTTAAGA	1239
Qy	101	AapProPothrA1a1yVa1a1aVa1G1uTrPheapBpG1yLybAapPheGInG1ySer	120
Db	1240	GACCAACCACTGGCCAGAGCTGCGCTGGAAATGTTGATGGAAAGATTTTCAGAGGAGC	1299
Qy	121	LybLeuLybVa1Ser1eua1aArg1yLybProPomeAaenSerMe1arG1yG1yLeu	140
Db	1300	AAACTTAAGTCTCCCTTGCTCGAGAAAGCCTCCAAATGAACAGATATCGGGGTGTCTG	1359
Qy	141	ProProAarg1uG1yAArg1yMeCProProPoleuAarg1yG1yProG1yG1yProG1y	160
Db	1360	CCACCCCTGAGGGGAGAGGACATGCAACACACTCCTGTAGAGTCCAGAGGCCACAGA	1419
Qy	161	G1yProG1yG1yProMeG1yAArgMeG1yG1yAArgG1yG1yAaPaArgG1yG1yPbaPro	180
Db	1420	GGTCTGGGGGACCATGGTGTGCAATGGAGAGCCGTGGAGAGATAGAGAGCTTCCCT	1479
Qy	181	ProAarg1yProAarg1ySerAarg1yAaNPoSerG1yG1yAaenVa1G1nH1aArg	200
Db	1480	CCAAGAGGACCCCGGGGTTCCTCGAGAGGAACCTCTGGAGGAGGAACGTCAACACGA	1539
Qy	201	AlaG1yAapTrpG1nCynProAaNPProG1yCybG1yAaenG1naaNPheA1aTrpAArgThr	220
Db	1540	GCTGGAGCTGACAGTGTCCCAATCCGGGTGTGGAAACAGAACTTCGCTGGAGAGCA	1599
Qy	221	G1uCybAaenG1nCylbA1aProLybProG1uG1yPheLeuProProPheProPro	240
Db	1600	GAGTGCACACAGTGTAAAGCCCCCAAGCCTBAAGCTTCTCCGCCACCTTCCGCC	1659
Qy	241	ProG1yG1yAaPaArg1yAArg1yG1yProG1yG1yMe1arG1yG1yAArg1yG1yLeu	260
Db	1660	CCGGGTGTGTATCGTGGCAGAGGTGGCCCTGTGTGACATGCGGGAGGAAAGGTGGCTTC	1719
Qy	261	Me1aPaArg1yG1yProG1yG1yMeCPhaArg1yG1yAArgG1yG1yAaPaArg1yG1y	280
Db	1720	ATGATCTGTGTGTGTCCTCGGTGAATGTTCAGAGGTGCGCTGTGGAGAGACAGAGTGGC	1779
Qy	281	PhaArg1yG1yAArg1yMeCAsPaArg1yG1yPheG1yG1yAArgG1yG1yAArgG1yPro	300
Db	1780	TTCCGTGTGTGCGGGGAGCATGACAGAGTGGCTTGTGTGAGGAAGCAGAGTGGCCT	1839
Qy	301	G1yG1yProProG1yProLeuMeG1uG1nMeTGTyG1yAArgAArg1yG1yAArg1yG1y	320
Db	1840	GGGGGGCCCCCTTGACCTTTATGTGAACAGATGGGAGAGAMAGAGGAGGACCTGGAGGA	1899
Qy	321	ProG1yLybMeCAsPlyG1yG1nH1aArg1nG1uTrgAArgAaPaArgProTyR	338
Db	1900	CTGGAATAATGTATTAAGGCGAGCACCTGTAGAGGCGCAGAGATCGGCCCTTAC	1953

RESULT 8	
CR610888	
LOCUS	CR610888
DEFINITION	2126 bp mRNA linear HTC 21-JUL-2004
	full-length cDNA clone CSDDF037Y107 of fetal brain of Homo sapiens
ACCESSION	CR610888
VERSION	CR610888.1 GI:50491695
KEYWORDS	HTC; CNSLT_CDNA.
SOURCE	Homo sapiens (human)
ORGANISM	Homo sapiens.

REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE	1 (bases 1 to 2126)
JOURNAL	Ll,W.B., Gruber,C., Jeesee,J. and Polayes,D. Full-length cDNA libraries and normalization Unpublished
REMARK	Contact : Feng Liang Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com/ Invitrogen Corporation 1600 Radway Avenue 2 (bases 1 to 2126) Genoscope. Direct Submission Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage : BP 191 91006 Evry cedex - FRANCE (E-mail : secrete@genoscope.cns.fr - Web : www.genoscope.cns.fr) 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five primers end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies, a division of invitrogen.
FEATURES	Location/Qualifiers
source	1..2126 /organism="Homo sapiens" /mol_type="mRNA" /db_xref="taxon:9606" /clone="CSODP037Y107" /tissue type="Fetal brain" /plasmid="pCMVSPORT_6"
ORIGIN	
Alignment Scores:	
Pred. No.:	3.34e-108 Length: 2126
Score:	1907.50 Matches: 337
Percent Similarity:	99.70% Conservative: 0
Best Local Similarity:	99.70% Mismatches: 0
Query Match:	99.25% Indels: 1
DB:	3 Gaps: 1
US-10-791-017A-2_COPY_319_656 (1-338) x CR610888 (1-2126)	
Oy	1 G ggllyatroglyygmctgyseralaglglnarglyglypheasnlvproglyy 20
Dd	943 GGAGGACCGGGGTGAATGGGC--GCTGGAGAGGAGAGGTGGCTCAATAAGCCGTGGTGA 999
Oy	21 PrometAspgluglyProAspLeuaspLeuglyProProValaaPaoAspgLuaspSer 40
Dd	1000 CCCATGGATGAAGAACCAGACTTTGATCTGAGGCCCACTGTAGATCCAAGTAGAACTCT 1059
Oy	41 AspaAnseralailetyrvalingilyleuasnaAspserValThrlieAspAspLeuala 60
Dd	1060 GACAACAGTGCAATTATGTACAAAGATTAATAATGACAGTGTGACTTATGATGATCTGGCA 1119
Oy	61 AsphepbelysgnCyeglyvaVallyMetasnlysaArgthrlyglinPrometlle 80
Dd	1120 GACTTCCTTTAAGCAGGTGGGGGTGTTAAGATGMAACAAGAACTGGCAACCATATATC 1179
Oy	81 HisllerytleuaaplysgluThrghlyvsProbglyAspalatrValaserTyrglu 100
Dd	1180 CACATCTAACCTCGGCAAGAAACGAAAAGCCCCAAGGCCATGCCACAGTGTCTTATGAA 1239
Oy	101 AspProothralalyalaialavalglutrpheaapglylysaapPheginglyser 120
Dd	1240 GACCACCACTGCCAAGGGGTGCCGTGGAAATGGTTGATGGGAAAAAGATTTTCAAAGGAGC 1299
Oy	121 LyaleulysvAlserleualahgylvlyysProPromeAnsersmetArnglylyleu 140
Dd	1300 AAATCTTAAAGCTCCCTTGTCCGGAAMAAGCTTCAATGAACGATATGCGGGGTGCTG 1359
Oy	141 ProProarogluglyAtarglyymeProProProleuanrglyglyProglylygProgly 160
Dd	1360 CCACCCCCTGATGGGCAAGGAGGATCCACCACTCCGTGTGAAGGTCCAGAGGCCCCAGGA 1419
Oy	161 GllyProglylyPrometgLyargMeGelyglyArgglylyAsparGglylyPhePro 180


```

Db      1420 GGTCTGGGGAGCCATGGGTGGCATGGAGGCGCGTGGAGAGATGAGAGGCTTCCTCCT 1479
Qy      181 ProArgGlyProArgGlySerArgGlyValanProSerGlyGlyValanValGlnHisArg 200
Db      1480 CCAAGAGAGACCCCGGGGTTCCCGAGAGAACCCCTCTGGAGAGAGAAACGTCGACACCA 1539
Qy      201 AlaGlyAspTrogInCySpProAnProGlyCyGlyValanGlnAnspPheAlaTTPArgThr 220
Db      1540 GCTGGAGACTGGACAGTGTCCCATCCGGGTTGTGGAAACCAAACTTCGCTGGAAGACA 1599
Qy      222 GluCyAsnGlnCyAlaValaProLysProGlyGlyPheLeuProProPhePro 240
Db      1600 GAGTCGAACCAAGTAAAGGCCCAAGAGCTTCCTCCGACACCTTCCTCCGACC 1659
Qy      241 ProGlyValAspArgGlyValArgGlyValProGlyValMetArgGlyValArgGlyValLeu 260
Db      1660 CCGGGTGTGTATGTGGCAGAGGTGGCCCTGTGGCAGTGGGAGAGAGAGAGGTGGCTTC 1719
Qy      261 MetAspArgGlyValProGlyValMetPheArgGlyValArgGlyValAspArgGlyVal 280
Db      1720 ATGATCGTGTGTGTCCCGGTGAATGTTCAAGGTGTGCGGTGTGGAGACAGAGGTGGC 1779
Qy      281 PheArgGlyValArgGlyValMetAspArgGlyValPheGlyValValArgArgGlyValPro 300
Db      1780 TTCGTGTGTGGCCGGGACATGACCGAGGTGGCTTTGTGTGAGAGAGACGAGGTGGCTTC 1839
Qy      301 GlyValProProGlyProLeuMetGlnMetGlyValArgArgGlyValArgGlyVal 320
Db      1840 GGGGGGCCCCCTGACCTTTGATGAGAACAGATGGAGAGAGAGAGAGAGAGAGAGAGAGAG 1899
Qy      321 ProGlyValMetAspLysGlyGlnHisArgGlnGlnArgArgAspArgProTyr 338
Db      1900 CTGGAAATAATGATTAAGCGAGACACCTCTCAGAGCGCAGAGATCGGCTTAC 1953

RESULT 9
CR625247 2159 bp mRNA linear HTC 21-JUL-2004
LOCUS CR625247
DEFINITION full-length cDNA clone CS0DB05F11 of Placenta of Homo sapiens
(human).
ACCESSION CR625247 GI:50506054
VERSION CR625247.1
KEYWORDS HTC; CNSLT cDNA.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 2159)
AUTHORS Li, W. B., Gruber, C., Jesse, J. and Polayes, D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished
REMARK Contact : Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/Invitrogen Corporation 1600
Faraday Avenue
2 (bases 1 to 2159)
REFERENCE Genoscope.
AUTHORS Direct Submission
TITLE Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage :
JOURNAL BP 131 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
COMMENT 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
was normalized. Library was constructed by Life Technologies, a
division of Invitrogen.
FEATURES
Location/Qualifiers
1..2159
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon.9606"
/clone="CS0DB05F11"
/issue_type="Placenta"
/plasmid="pCMVSPORT_6"

```

```

ORIGIN
Alignment Scores:
pred. No.: 3 4e-108 Length: 2159
Score: 1907.50 Matches: 337
Percent Similarity: 99.70% Conservative: 0
Best Local Similarity: 99.70% Mismatches: 0
Query Match: 99.25% Indels: 1
DB: 3 Gaps: 1

US-10-791-017a-2_COPY_319_656 (1-338) x CR625247 (1-2159)
Qy      1 GlyIArgGlyValMetGlySerAlaGlyValArgGlyValPheAnLysProGlyVal 20
Db      1000 GGAGAGCGGGTGGATGGGC---CGTGAAGAGCGAGGTGGCTTCAATTAAGCTGTGTGA 1056
Qy      21 PrometAspGlnGlyProAspLeuAspLeuGlyProProValAspProAspGluAspSer 40
Db      1057 CCCATGATGAAGAGACCAATCTTGTATCTAAGGCCCACTGTAGATCCAGATGAAGACTCT 1116
Qy      41 AspAsnSerAlaIleTyrValGlnGlyLeuAsnAspSerValThrLeuAspLeuAla 60
Db      1117 GACACAGTCGACATTTATATGACAGATTAAATGACAGTGTACTTATGATGATCTGGCA 1176
Qy      61 AspPhePheLysGlnCyGlyValValLysMetAsnLysArgThrGlyGlnProMetIle 80
Db      1177 GACTCTTAAACAGTGTGGGTTGTTAAGATGAACAAAGAACTGGCAACCATGATTC 1236
Qy      81 HisIleTyrLeuAspLysGlnThrGlyLysProLysGlyValAspAlaThrValSerTyrGln 100
Db      1237 CACATCTTACCTGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1296
Qy      101 AspProProThrAlaLysAlaValAlaValGlnTyrPheAspGlyValAspPheGlnGlySer 120
Db      1297 GACCAACCACTGCCCAAGCTGCGGTGGAAATGTTGATGGAGAAATTTCAAGGAGAC 1356
Qy      121 LysLeuLysValSerLeuAlaArgLysLysProPheMetAsnSerMetArgGlyValLeu 140
Db      1357 AAACCTTAAGTCTCCCTTCTCGAAGAGAGCTCCCAATGAACAGTATCGGGGTGTCTTG 1416
Qy      141 ProProArgGlnGlyValArgGlyValMetProProProLeuArgGlyValProGlyVal 160
Db      1417 CCAACCCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1476
Qy      161 GlyProGlyValProMetGlyValArgMetGlyValArgGlyValValAspArgGlyValPhePro 180
Db      1477 GGTCTGGGGAGCCATGGGTGGCATGGAGGCCGTGGAGAGATGAGAGAGCTTCCTCCT 1536
Qy      181 ProArgGlyProArgGlySerArgGlyValanProSerGlyGlyValanValGlnHisArg 200
Db      1537 CCAAGAGAGACCCCGGGGTTCCGAGAGGAACCCCTCTGAGAGAGAGAAACGTCGACACCA 1596
Qy      201 AlaGlyAspTrogInCySpProAnProGlyCyGlyValanGlnAnspPheAlaTTPArgThr 220
Db      1597 GCTGGAGACTGGACAGTGTCCCAATCCGGGTGTGGAAACCAAACTTCGCTGAGAGACA 1656
Qy      221 GluCyAsnGlnCyAlaValaProLysProGlyGlyPheLeuProProPhePro 240
Db      1657 GAGTCGAACCAAGTAAAGGCCCAAGAGCTTCCTCCGACACCTTCCTCCGACC 1716
Qy      241 ProGlyValAspArgGlyValArgGlyValProGlyValMetArgGlyValArgGlyValLeu 260
Db      1717 CCGGGTGTGTATGTGGCAGAGGTGGCCCTGTGGCAGTGGGAGAGAGAGAGAGAGAGAGAG 1776
Qy      261 MetAspArgGlyValProGlyValMetPheArgGlyValArgGlyValLysAspArgGlyVal 280
Db      1777 ATGATCGTGTGTGTCCCGGTGAATGTTCAAGGTGTGCGGTGTGGAGACAGAGGTGGC 1836
Qy      281 PheArgGlyValArgGlyValMetAspArgGlyValPheGlyValValArgArgGlyValPro 300
Db      1837 TTCGTGTGTGGCCGGGACATGACCGAGGTGGCTTTGTGTGAGAGAGAGAGAGAGAGAGAG 1896
Qy      301 GlyValProProGlyProLeuMetGlnMetGlyValArgArgGlyValArgGlyVal 320

```


TITLE	Okazaki, Y., Muramatsu, M., Inoue, Y., Kitz, A. and Hayashizaki, Y. RIKEN Integrated sequence analysis (RISA) system-384 format Benchmarking pipeline with 384 multicapillary sequencer Genome Res. 10 (11), 1757-1771 (2000)
JOURNAL	20530913
MEDLINE	11076861
REFERENCE	
AUTHORS	4
TITLE	The RIKEN Genome Exploration Research Group Phase II Team and the PANTOM Consortium. Functional annotation of a full-length mouse cDNA collection Nature 409, 685-690 (2001)
JOURNAL	5
REFERENCE	
AUTHORS	
TITLE	Group Phase I & II Team. Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs Nature 420, 563-573 (2002)
JOURNAL	6 (b) (1 to 2373)
REFERENCE	
AUTHORS	Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hirooka, T., Hirozane, T., Hori, F., Imocani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Kato, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohashi, N., Okazaki, Y., Saito, K., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shingawa, A., Shiraki, T., Sogabe, Y., Tsgami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toy, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.
TITLE	Direct Submission
JOURNAL	Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gs.c.riken.jp, URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)
COMMENT	cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. Please visit our web site for further details. URL: http://genome.gsc.riken.jp/ URL: http://pantom.gsc.riken.jp/ Location/Qualifiers
FEATURES	1..2373 location=Qualifiers /organism="Mus musculus" /mol_type="mRNA" /strain="C57BL/6J" /db_xref="PANTOM:DB:9430031M11" /db_xref="taxon:10090" /clone="9430031M11" /tissue_type="embryonic body between diaphragm region and neck" /clone_11b="RIKEN full-length enriched mouse cDNA library" /dev_stage="12 days embryo" 43..2011 /note="Bwing sarcoma homolog (MCD MGI:99960, GB NM_007968 evidence: BLASTN, 99%, match=2172) putative" 2358..2363 /note="putative" 2373 /note="putative" /note="putative"
ORIGIN	
Alignment Scores:	6.36e-106 Length: 2373 Pred. No.: 1871.50 Matches: 330 Score: 98.82% Conservative: 4 Percent Similarity: 97.63% Mismatches: 3 Best Local Similarity:

Query Match:	97.37%	Indels:	1
DB:	3	Gaps:	1
US-10-791-017A-2_COPY_319_656 (1-338) x AK034755 (1-2373)			
QY 1	G1yG1yArGg1yG1yMeG1ySeRa1aG1yG1yAaRg1yG1yPheAen1ySProG1yG1y		20
Db 998	GGAGGAGCGCGGTGGACTGGCG---GCTGGAGAGCGAGGTGGCTTCAATAAGCCTGGTGA		1054
QY 21	PrometAaPgi1uG1yProAaPleuAaPleuG1yProProV1A1AaPProAaPgi1uAaPSeR		40
Db 1055	CCCAATGGATGAAGGACCAAGATCTTGAATCTTAAGGCTTCTTAATGAATCCCAATGAAGACTCT		1114
QY 41	AaPaenSeRa1a1e1yVa1G1nG1y1LeuAaPaAaPSeR1a1Th1eAaPaA1e1u1A		60
Db 1115	GACAAAGGTCAATTTATGTGCAAGATTTAAATGACATGTGACTCTGATGATGTGCA		1174
QY 61	AaPheAe1ySg1yCg1yVa1Va1yMeAaSn1ySaRgThRg1yG1nProMeT1e		80
Db 1175	GACTTCTTTAAGCAGTGTGGGTGTGTCAATGATGAACAAGAACTGGACAACCATGATC		1234
QY 81	Hi1e1e1yR1e1uAaP1ySg1yThRg1y1ySPro1ySg1yAaP1a1ThR1e1S1e1yG1u		100
Db 1235	CATATCTCACTCGATTAAGAGACAGAAACCTTAAGAGGACGCAAGTCTTAAGAA		1294
QY 101	AaPProThR1a1ySa1a1a1a1a1G1uTTPheAaP1y1ySaAaPheG1nG1ySeR		120
Db 1295	GATCCACCAACTGCAAAAGCTGCCGTGGATGTGTTATGGAAAGTTTCAAGAAAGC		1354
QY 121	1yA1e1u1yA1S1e1u1a1a1rG1yA1ySProPProMeAaSeR1eAaRg1yG1y1e1u		140
Db 1355	AAACTTAAGTGTCTTCCCGCAAGAAACCTTCAATGAACAGACGCGGAGGCGATG		1414
QY 141	ProProAaRg1uG1yAaRg1yMeProProPro1e1uAaRg1yG1yProG1yG1yProG1y		160
Db 1415	CCACCTGTGAAGGCAAGAGAGATCCACCACTTGTGAAGTCTGTGGTCCAGGA		1474
QY 161	G1yProG1yG1yProMeTg1yAaRmeTg1yG1yAaRg1yG1yAaPaRg1yG1yPhePro		180
Db 1475	GGCCCTGAAGGACCAATGGGTGGATGGAGGCGCGTGAAGAGACAAAGGGGGTTCCT		1534
QY 181	ProAaRg1yProAaRg1ySeRaRg1yAaSnProSeRg1yG1yAaSnVa1G1nH1aRg		200
Db 1535	CCAAGAGGCGCCCGAGGCTCCAGAGGAACCCCTCTGAAGAGGAATGTGCCAGACCGA		1594
QY 201	AlaG1yAaPTrpG1nCySProAaPProG1yCyG1yAaSnG1nAaSnPheA1aTPPaRgThR		220
Db 1595	GCTGAAGATCGGCAAGTGTCCCAATCAAGAGGCTGTGAAGAACCAAGATCTTGAGAAACA		1654
QY 221	G1uCySaAaSnG1nCyS1ySa1a1aPro1yAaPProG1uG1yPhe1uAaPProProPhePro		240
Db 1655	GAAATGCAACAGTGTAAAGGCCCTTAACCCGAGGGCTTCTCCCGGCAACCTTCCAACT		1714
QY 241	ProG1yG1yAaPaRg1yAaRg1yG1yProG1yG1yMeCaRg1yG1yAaRg1yG1y1e1u		260
Db 1715	CCGGGTGTGTCTGTGAACGAAGTGGCGCTGTGGCAATGCAAGAGAAAGGAGGACTC		1774
QY 261	MeT1eAaPaRg1yG1yProG1yG1yMeT1eAaRg1yG1yAaRg1yG1yAaPaRg1yG1y		280
Db 1775	ATGAACCGTGTGTCTCTGAAGAAATGTCAAGAGTGTCAAGAGTGTGAAGACAGAGAGGC		1834
QY 281	PheAaRg1yG1yAaRg1yMeT1eAaPaRg1yG1yPheG1yG1yG1yAaRg1yG1yPro		300
Db 1835	TTCCGAAGTGGCCGTGAATGTACGAAGTGGCTTTGTGTGAAGGAAGCGAGTGTCT		1894
QY 301	G1yG1yProProG1yPro1e1uMeT1uG1nMeT1yG1yAaRg1yG1yAaRg1yG1y		320
Db 1895	GAGGAGGCTCTCTGTGAACCTTTAATGAACAGATGGAGAAAGAAAGCGGCACTGGAGGA		1954
QY 321	ProG1y1yMeT1eAaR1yG1yG1nH1aRg1G1nG1yAaPaRPaRPaRPaRPaRPro1yR		338
Db 1955	CTGTGAGAAATGAATTAAGGAGACACGTCGTAGGAACGCAAGAGCCGGCCTTAC		2008

RESULT 13
CR604639
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
REMARK
REFERENCE
AUTHORS
TITLE
JOURNAL
REMARK
COMMENT
FEATURES
source
ORIGIN
Alignment Scores:
Pred. No.:
Score:
Percent Similarity:
Best Local Similarity:
Query Match:
DB:
US-10-791-017a-2_COPY_319_656 (1-338) x CR604639 (1-2236)

CR604639 2236 bp mRNA linear HTC 21-JUL-2004
full-length cDNA clone CS0DU013YN08 of T cells (Jurkat cell line)
Cot 10-normalized of Homo sapiens (human).
CR604639
CR604639.1 GI:50485446
HTC; cNSLT cDNA.
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 2236)
L.J.W.B., Gruber, C., Jessee, J. and Polyes, D.
Full-length cDNA libraries and normalization
Unpublished
Contact: Feng Liang Email: fliang@lifetech.com URL:
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Faraday Avenue
2 (bases 1 to 2236)
Genoscope.
Direct Submision
Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage:
BP 191 91006 EVRY cedex - FRANCE (E-mail: seqref@genoscope.cns.fr
- Web: www.genoscope.cns.fr)
1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
end enriched, double-strand cDNA was digested with Not I and cloned
into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
was normalized. Library was constructed by Life Technologies, a
division of Invitrogen.
Location/Qualifiers
1..2236
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DU013YN08"
/issue_type="T cells (Jurkat cell line) Cot
10-normalized"
/plasmid="pCMVSPORT_6"

4,32e-98 Length: 2236
1744.50 Matches: 314
92.90% Conservative: 0
92.90% Mismatches: 1
90.76% Indels: 24
Gaps: 2

1 G1G1YARGLYGLYMETGLYSERIALAG1YGLUARG1YGLYPHEANLYSPROGLYGLY 20
973 GGAAGACGCGGTGGATGGGC---GCTGAGAGCGAGGTGGCTTCAATTAAGCTGTGGA 1029
21 PROMETASPG1UGLYPROASPLEUAPLEUGLYPROPOVALASPPROASPG1UASPSER 40
1030 CCCATGATGAGAGACCAATCTTGTATCTAGGCCCACTGTAGATCCAGATGAAGACTCT 1089
41 AAPAASERLALIEYRVAG1GNG1YLEUAASPSERVALTHREUASPSPLEUALA 60
1090 GACCAACAGTCAATTTATATACAGGATTAAGAGAGAGTGTACTTACATGATGATCGGA 1149
61 AAPPEPHELYESGLNCGYAG1YVAL1YBMEASPLNLYARGTHRG1YGLINPROMETILE 80
1150 GACTTCTTTAAGCAGTGGGTGTGTTAAGATGACAAAGAACTGGGCAACCATGATC 1209
81 HSEILTYRLEUASPLYSGLUTRG1YLYSPROLYSG1YASPALATHRYALISERTYGLU 100
1210 CACATCTACTGTGACAAAGAAACAGGAAAGCCCAAGGCGATGCCACAGTGTCTATGAA 1269
101 AAPPPROCTHTRALALYVALA1AVALG1UTTPHESAG1YLYEASPSHEGLNGLYSER 120

Db 1270 GACCCACCACTGCCAAGGTGCGGTGGATGGTTTGATGGGAAAGTTTCAAGGAGC 1329
QY 121 LYELEULYVALSERLEUALAARG1YLYSPROPMETASPSERMETARG1YGLYLEU 140
Db 1330 AAACCTTAAGTCTCCCTTGTCTCGAAGAACCTCCATGAACAGTATGCGGGGTGGCTG 1389
QY 141 PROPMARG1UGLYARG1YMETPROPMOPROPMLEUARG1YGLYPROGLY 160
Db 1390 CCACCCGTGAGGACAGGACATGACCACTCCGTGAGAGTCCAGAGGCCACGA 1449
QY 161 GLYPROGLYGLYPROMETGLYARGMETGLYGLYARG1YGLYASPARGLYGLYPHEPRO 180
Db 1450 GGTCTGGGGGACCATGGGTGGCATGGAGGCCGTGGAGAGATAGAGAGCTTCCCT 1509
QY 181 PROARG1YPROARG1YSEARG1YASMPROSEARG1YGLYASVALG1NH1SARG 200
Db 1510 CCNAGAGACCCCGGGGTTCCCGAGGAAACCCCTTGAGAGAGAAACGTCAGACCGA 1569
QY 201 ALAAGLYASPTRG1NCGYSLYASPSANPARG1YCYG1YASNG1YASNPHEALATPARGTHR 220
Db 1570 GCTGAGACTGGCAGTGTCCCAATCCGGGTGTGGAAACCAACTTCGCTTGAGACCA 1629
QY 221 GLUCYASNG1NCGYSLYASPSANPARG1YCYG1YASNG1YASNPHEALATPARGTHR 240
Db 1630 GAGTGCACCACTGTGATGAGCCCAAGCTTGAAGGCTTCTCCGCGCACCTTTCCGCC 1689
QY 241 PROGLYGLYASPARGLYARG1YGLYPROGLYGLYMETARG1YGLYARG1YGLYLEU 260
Db 1690 CCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1749
QY 261 METASPARGLYGLYPROGLYGLYMETPHEARG1YGLYARG1YGLYASPARGLYGLY 280
Db 1750 ATGGAATCGTGTGTGTCCCGGTGAATGTTCAGAGGTGGCGGTGTGAGACAGAGGTGGC 1809
QY 281 PHEARG1YGLYARG1YMETASPARGLYGLYPHEGLYGLYGLYARG1YGLYARG1YGLY 300
Db 1810 TTCGGTGTGGCGGGGCAATGACGAGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1869
QY 301 G1YGLYPROGLYPROLEUWETGLUG1METGLYGLYARG1YGLYARG1YGLY 320
Db 1870 GGGGGGCC-CCA----- 1880
QY 321 PROGLYLYMETASPLYSGLYLUHISARG1GLUARG1YARG1YGLYARG1YGLY 338
Db 1881 -----GGCGAGCACCTCAGAGCGAGAGATCGCCCTAC 1916
RESULT 14
AK049743 2269 bp mRNA linear HTC 03-APR-2004
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED

AK049743 2269 bp mRNA linear HTC 03-APR-2004
Mus musculus 12 days embryo spinal cord cDNA, RIKEN full-length
enriched library, clone: C530046A18 product: bwing sarcoma homolog,
full insert sequence.
AK049743
AK049743.1 GI:26093614
HTC; CAP Trapper.
Mus musculus (house mouse)
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sclirognathi; Muridae; Murinae; Mus.
1 Carninci, P., and Hayashizaki, Y.
High-efficiency full-length cDNA cloning
Meth. Enzymol. 303, 19-44 (1999)
99279253
10349636
2 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes
Genome Res. 10 (10), 1617-1630 (2000)
JOURNAL
MEDLINE
PUBMED
20499374
11042159

Db	1754	GGTGGCTTTGGTGGAGGAAGACGAGGTGTCTCTGGGGGGCCCTTCGACCTTTATGSA	1813
Qy	310	GINmetcglvgyIvArxArGIvgyIvArxgIvgyIvProcIvlyvMetArbPlygIvGIvHb	329
Db	1814	CAGATGGGAGGAAAGAGCGCGACGTGGAGGACCTGGGAAATGGATTAAGGCGAGCAC	1873
Qy	330	ArGcInGIvArxArGArGArGArGProtyr	338
Db	1874	CGTCAGGAACGACAGACCGCGCCCTTAC	1900
RESULT 15			
HSMB04802			
LOCUS	HSMB04802	4729 bp	mRNA linear HTC 03-AUG-2004
DEFINITION	Homoe sapiens mRNA, CDNA DKFZp686i2029 (from clone DKFZp686i2029).		
ACCESSION	AL833489		
VERSION	AL833489.1	GI:21734132	
KEYWORDS	HTC.		
SOURCE	Homoe sapiens (human)		
ORGANISM	Homoe sapiens		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
REFERENCE	1 (bases 1 to 4729)		
AUTHORS	Bahr, A., Ladoer, J., Mewes, H.W., Weill, B., Amid, C., Osanger, A.,		
	Pobo, G., Han, M. and Wiemann, S.		
CONSRM	The German CDNA Consortium		
TITLE	Direct Submission		
JOURNAL	Submitted (13-JUL-2004) MIPS, Ingolstaedter Landstr.1, D-85764		
COMMENT	Nueberberg, GERMANY		
	Clone from S. Wiemann, Molecular Genome Analysis, German Cancer		
	Research Center (DKFZ), Email s.wiemann@dkfz-heidelberg.de;		
	Sequenced by Qiagen (Hilden/Germany) within the CDNA sequencing		
	consortium of the German Genome Project.		
	This clone (DKFZp686i2029) is available at the RZPD Deutsches		
	Reesourcenzentrum fuer Genomforschung GmbH in Berlin, Germany.		
	Please contact RZPD for ordering:		
	http://www.rzpd.de/cgi-bin/products/cl.cgi?CloneID=DKFZp686i2029		
	Further information about the clone and the sequencing project is		
	available at http://mips.gsf.de/projects/cdna/.		

FEATURES

```

1. .4729
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="RZPD:DKFZp686I2029"
/db_xref="taxon:9606"
/clone="DKFZp686I2029"
/tissue_type="testis"
/clone_lib="666 (synonym: hlec3). Vector pSport1_srl; host
DH10B; sites SfiI + SfiIb"
/dev_stage="adult"
/notes="RNA-binding protein EMS, N-terminus truncated, not
fully spliced"

```

ORIGIN

Alignment Scores:	
Pred. No.:	1.81e-93
Score:	1675.50
Percent Similarity:	61.04%
Best Local Similarity:	61.04%
Query Match:	87.17%
Length:	4729
Matches:	318
Conservative:	0
Mismatches:	1
Indels:	203

US-10-791-017A-2_COPY_319_656 (1-338) x HSM804802 (1-4729)

QY 20 G L Y P r o m e t A P G L I n g L Y P r o A s p l e u n s p l e u g l Y P r o C o V a L l a s P r o A s p l e u A s p 39
D b 2964 G S A C C C A T G C A T G A N G A C C A G A C T T T A T C T A G C C C A C C T G T A G A T C C A G A T G A A G A C 3023
QY 40 S e r A p a n s e r r a l l e t y r V a l G l n g l Y l e u n s p S e r v a l T h r l e u A s p A s p l e u 59
D b 3024 T C T G C A C A C A G T C A T T A T G T A C A A G A T T A A T G A C A G T G A C A C T C T G A T G A T C T G 3083
QY 60 A l A a p p h e p h e l y P G l n C y a g l Y v a l V a l l y s h e t a n l y s A a g t t h r g l Y g l n P r o m e t 79

Db	3084	GCAGACTCTTTAAGCACTGCGGGTTCTTAAGTAAAGAAACAAGAAACTGGCAACCATG	3143
Qy	80	IllehisileIyRLeuaspLygIuThrGlyIysPProLygIyAspAlaThrValSerTy	99
Db	3144	ATCCACATCTACCTGGACAAAGAAACAGAAAGCCCAAGGCGATGCCACAGTCTCCAT	3203
Qy	100	GIuAsPProPothrAlaIysAlaIValGIuItrPheAspGlyIyAspPhgIuGI	119
Db	3204	GAAGACCAACCACTGCACAGGCTGCCTGGAAATGTTGATGGAAAGATTTCAAGGG	3263
Qy	120	SerLySLeuLyValSerLeuAlaArgLySAsPProPromeLanSerMetArgIyGI	139
Db	3264	AGCAAACTTAAAGTCTCCCTTGCTCGAGAAAGCCTCCATGAACAGTATCGGGGTGT	3323
Qy	140	LeuPProPArgIuGIyArgIyMecPProPProLeuArgIyGIyProGIyGIPro	159
Db	3324	CTGCCACCCCGTGAAGGCGAGAGGATGCCAACCACTCCGTGAGGTCCAGAGGCCCA	3383
Qy	160	GIyGIyPProGIyGIyPProMetGIyArgMetGIyGIyArgGIyGIyAspArgGIyGIyPhe	179
Db	3384	GGAGGTCTGGGGGAGCCATGGGTGCATYGGAGGCCGTGGAGAGATPAGAGAGGCTTC	3443
Qy	180	ProPProArgIyPProArgIySerArgGIyAsnProSerGIyGIyIysAsnValGIuHis	199
Db	3444	CCTCCAAAGAACCCCGGGGTTCCCGAGGAGACCCCTCTGAGAGGAACGTCCAGCAC	3503
Qy	200	ArgAlaGIyAsPTrPIuGIyCysPProAsnProGIy-----	210
Db	3504	CGAGTGGAGACTGGCAGTGTCCCAATCCGA-TGTACTGTCTTGGCAAAATTGATACC	3562
Qy	210	-----	210
Db	3563	TACGATGAAGCAACCTTCCCTCACCCCATCCCACTAGAGTGATTGCTGTCTA	3622
Qy	210	-----	210
Db	3623	GAGGAACAAATGATGACCTGATGGCTGGTTAGGAACTAGTCAGCAATTCACTGAC	3682
Qy	210	-----	210
Db	3683	GCTTCAGAGCTTCTGAAGATTGATTGACCTGTCTGTGGGTGCATGCTGCTGAGGC	3742
Qy	210	-----	210
Db	3743	TGTGCCCTAAAGCATGGGTGTACATAGACTCTTGATAGAGTGTACTCTGTACA	3802
Qy	210	-----	210
Db	3803	CACCACTTCTCTGTATTCTCTCTTAAGTTCAATTGTGATTTCTGTGATGTAAAT	3862
Qy	211	-----CysGIyAsnGIuAsnPhelAtPTrArgThrGIuCysAsnGIuCysLyS	226
Db	3863	GTATGCAAGGGTGTGTGAACCAACAACTTGCCTCGAGAAACAGTGTCAACACTGTAAAG	3922
Qy	227	AlaPProLySPProGIuGIyPheLeuPProPProPhePProPPro-----	241
Db	3923	GCCCCAAGCTGAAGGCTTCTCTCCGCCACCTTTCTCGCCCCGGGTAGTGCAGGTTT	3982
Qy	241	-----	241
Db	3983	CATGATGTCCCTCAGCTTCTCTGTGCTAAACCTCTTTCTTAATTTGTGGCTGTGTAA	4042
Qy	241	-----	241
Db	4043	ACTGCAGTGCCCTCTGTCTTAACAACATTGATGTGTGTCTCATTTCTAAATTGTCA	4102
Qy	241	-----	241
Db	4103	GCCGATGCCGAGATTGATGAAGTGTCTGTGTTTGTCTGTGTGAGAGAAGAACAGA	4162
Qy	241	-----	241
Db	4163	GCAGCTTCAAGTGTCAACAGGCGCTTGCAGCAACCACTGACTGTTCGCTTGCCTGCT	4222

```
QY 242 -----GlyGlyAspArgGlyArgGlyGlyProGlyGlyMetArgGlyArg 257
Db 4223 ATTCTACCTTAGGTGTGATCGTGGCAGGTGGCCCTGTGTGCATGCCGGAGAGAG 4282
QY 258 GlyGlyLeuMetAspArgGlyGlyProGlyGlyMetPheArgGlyGlyArgGlyAsp 277
Db 4283 GTGGCCTCATGATGTGTGTGTCCTCGGTGGAATGTTCAAGAGTGGCCGTGTGGAGAC 4342
QY 278 ArgGlyGlyPheArgGlyGlyArgGlyMetAspArgGlyGlyPheGlyGlyArgArg 297
Db 4343 AGAGGTGGCTTCGTGTGTGCGCGGCATGACCGAGTGCTTTGTGTGAGGAGACGA 4402
QY 298 GlyGlyProGlyGlyProGlyProLeuMetGlyGlyMetGlyGlyArgArgGlyGly 317
Db 4403 GTGGCCCTGGGGGGCCCTGGACCTTTGATGACACAGATGGAGAGAGAGAGAG 4462
QY 318 ArgGlyGlyProGlyGlyMetAspGlyGlyGlyGlyHisArgGlyGlyArgArgAspArgPro 337
Db 4463 CGTGGAGGACCTGGAAAAATGATTAAGCGAGACACCGTCAAGAGCGCAGAGATCGGCC 4522
QY 338 Tyr 338
Db 4523 TAC 4525
```

Search completed: February 21, 2005, 04:12:05
Job time : 2560 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - nucleic search, using frame_plus.p2n model

Run on: February 21, 2005, 02:07:46 ; Search time 1870.22 Seconds
(without alignments)
1068.188 Million cell updates/sec

Title: US-10-791-017A-2_COPY_319_656
Perfect score: 1922
Sequence: 1 GCGKMGSGAGRGGRFKPGG.....GGRKADKGRHROBRDRP 338

Scoring table:
BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 5384158 seqs, 2955248155 residues

Total number of hits satisfying chosen parameters: 10768316

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Command line parameters:
-MODE=frame+ p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO.spool.p/US10791017/runat.17022005.125809.22196/app.query.fasta_1.1358
-DB=Published Applications NA -QFMT=fastbp -SUPERX=p2n.rnpb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bites -START=1 -END=1 -MATRIX=blonum62
-TRANS=human40.cdi -LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pico -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=200000000 -USER=US10791017.0CGN.1.1.879 @runat.17022005.125809.22196
-NCPU=6 -ICPU=3 -NO MAP -LARGESUBSTR -NEG SCORES=0 -WAIT -DSBLOCK=100
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-Fgapop=6 -Fgapext=7 -Ygapop=10 -Ygapext=0.5 -Delop=6 -Delext=7

Database : Published Applications_NA:
1: /cgn2_6/prodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/prodata/1/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/prodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/prodata/1/pubpna/US07_NEW_PUB.seq:*
5: /cgn2_6/prodata/1/pubpna/US07_PUBCOMB.seq:*
6: /cgn2_6/prodata/1/pubpna/US08_NEW_PUB.seq:*
7: /cgn2_6/prodata/1/pubpna/US08_PUBCOMB.seq:*
8: /cgn2_6/prodata/1/pubpna/US09_PUBCOMB.seq:*
9: /cgn2_6/prodata/1/pubpna/US09_PUBCOMB.seq:*
10: /cgn2_6/prodata/1/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/prodata/1/pubpna/US09_PUBCOMB.seq:*
12: /cgn2_6/prodata/1/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/prodata/1/pubpna/US10_PUBCOMB.seq:*
14: /cgn2_6/prodata/1/pubpna/US10_PUBCOMB.seq:*
15: /cgn2_6/prodata/1/pubpna/US10_PUBCOMB.seq:*
16: /cgn2_6/prodata/1/pubpna/US10_PUBCOMB.seq:*
17: /cgn2_6/prodata/1/pubpna/US10_PUBCOMB.seq:*
18: /cgn2_6/prodata/1/pubpna/US10_PUBCOMB.seq:*
19: /cgn2_6/prodata/1/pubpna/US10_NEW_PUB.seq:*
20: /cgn2_6/prodata/1/pubpna/US11_NEW_PUB.seq:*
21: /cgn2_6/prodata/1/pubpna/US60_NEW_PUB.seq:*
22: /cgn2_6/prodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1922	100.0	1988	17 US-10-094-749-1074	Sequence 1074, Ap
2	1922	100.0	2176	9 US-09-822-830A-49	Sequence 49, Appl
3	1922	100.0	2390	9 US-09-880-107-3769	Sequence 3769, Ap
4	1922	100.0	2390	10 US-09-860-706-1081	Sequence 1081, Ap
5	1922	100.0	2390	10 US-09-873-319-717	Sequence 717, Ap
6	1922	100.0	2390	18 US-10-791-017A-1	Sequence 1, Appl1
7	1777	92.5	2273	9 US-09-822-830A-410	Sequence 410, Appl
8	1181	61.4	2299	14 US-10-198-846-9847	Sequence 9847, Ap
9	961	50.0	550	10 US-09-918-995-27690	Sequence 27690, A
10	878	45.7	1822	18 US-10-755-889-649	Sequence 649, App
11	878	45.7	1824	17 US-10-439-703-58	Sequence 58, Appl
12	878	45.7	1939	10 US-09-919-039-322	Sequence 322, App
13	809.5	42.1	568	10 US-09-918-995-9557	Sequence 9557, Ap
14	718.5	37.4	545	9 US-09-925-101-669	Sequence 669, App
15	593	30.9	361	9 US-09-960-852-3851	Sequence 3851, Ap
16	493	25.7	503	9 US-09-864-761-8728	Sequence 8728, App
17	493	25.7	503	16 US-10-029-386-4463	Sequence 4463, Ap
18	486	25.3	262	16 US-10-029-386-18163	Sequence 18163, A
19	475	24.7	253	9 US-09-864-761-25438	Sequence 25438, A
20	460	23.9	457	18 US-10-469-285-488	Sequence 488, App
21	423	22.0	1401	18 US-10-437-963-8943	Sequence 8943, A
22	420	21.9	1397	17 US-10-425-114-31980	Sequence 31980, A
23	420	21.9	1397	18 US-10-425-115-42134	Sequence 42134, A
24	414	21.5	483	10 US-09-918-995-2865	Sequence 2865, Ap
25	391	20.3	1352	18 US-10-739-930-501	Sequence 501, Appl
26	364.5	19.0	1449	17 US-10-425-114-4694	Sequence 4694, Ap
27	364.5	19.0	1449	18 US-10-425-115-42131	Sequence 42131, A
28	364	18.9	1166	18 US-10-425-115-42135	Sequence 42135, A
29	344	17.9	450	9 US-09-738-973-145	Sequence 145, App
30	344	17.9	450	9 US-09-854-133-145	Sequence 145, App
31	344	17.9	450	15 US-10-144-649A-145	Sequence 145, App
32	319.5	16.6	1966	18 US-10-437-963-62027	Sequence 62027, A
33	314	16.3	321	9 US-09-933-797-351	Sequence 351, App
34	305	15.9	4428	17 US-10-402-089-11	Sequence 11, Appl
35	305	15.9	4428	17 US-10-402-089-3	Sequence 3, Appl1
36	303	15.8	4428	17 US-10-402-089-5	Sequence 5, Appl1
37	303	15.8	4428	17 US-10-402-089-5	Sequence 5, Appl1
38	303	15.8	4428	17 US-10-402-072A-3	Sequence 3, Appl1
39	303	15.8	4428	17 US-10-402-072A-5	Sequence 5, Appl1
40	301	15.7	5460	10 US-09-918-715-225	Sequence 225, App
41	301	15.7	5460	15 US-10-177-293-67	Sequence 67, Appl
42	301	15.7	5460	15 US-10-301-822-32	Sequence 32, Appl
43	301	15.7	5460	17 US-10-257-021-71	Sequence 71, Appl
44	301	15.7	5460	17 US-10-172-118-428	Sequence 428, App
45	301	15.7	5460	17 US-10-342-887-428	Sequence 428, App

ALIGNMENTS

RESULT 1
US-10-094-749-1074
; Sequence 1074, Application US/10094749
; Publication No. US20030219741A1
; GENERAL INFORMATION:
; APPLICANT: ISOGAI, TAKAO
; APPLICANT: SUGIYAMA, TOMOYASU
; APPLICANT: OTSUKI, TETSUJI
; APPLICANT: WAKAMATSU, AI
; APPLICANT: SATO, HIROYUKI
; APPLICANT: ISHII, SHIZUKO
; APPLICANT: YAMAMOTO, JUN-ICHI
; APPLICANT: ISONO, YUUKO
; APPLICANT: HIO, YUKI
; APPLICANT: OTSUKA, KAORU
; APPLICANT: NAGAI, KEIICHI
; APPLICANT: IRIE, RYOTARO
; APPLICANT: TAMECHIKA, ICHIRO
; APPLICANT: SEKI, MOTOHITO
; APPLICANT: YOSHIKAWA, TSUTOMU

APPLICANT: OTSUKA, MOTOKUKI
 APPLICANT: NAGAHART, KENJI
 APPLICANT: MASUHO, YASUHIKO
 TITLE OF INVENTION: NOVEL FULL-LENGTH cDNA
 FILE REFERENCE: 084335/0160
 CURRENT APPLICATION NUMBER: US/10/094,749
 CURRENT FILING DATE: 2002-03-12
 PRIOR APPLICATION NUMBER: 60/350,435
 PRIOR FILING DATE: 2002-01-24
 PRIOR APPLICATION NUMBER: JP 2001-328381
 PRIOR FILING DATE: 2001-09-14
 NUMBER OF SEQ ID NOS: 3381
 SOFTWARE: PatentIn Ver. 2.1
 SEQ ID NO 1074
 LENGTH: 1988
 TYPE: DNA
 ORGANISM: Homo sapiens
 US-10-094-749-1074

Alignment Scores:
 Pred. No.: 4e-155 Length: 1988
 Score: 1922.00 Matches: 338
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 17 Gaps: 0

US-10-791-017a-2_COPY_319_656 (1-338) x US-10-094-749-1074 (1-1988)

QY 1 G1YGYAARG1YGLYMETG1YSERA1AG1YGLUARG1YGLYPHEANLYSPROGLYGLY 20
 DB 798 GGAGGAGCGCGTGGAAATGGGAGCGCTGGAGAGAGAGTGGCTTCATTAAGCCGTGGGA 857
 QY 21 Prometaspgluclg1yproaspbleuaspleug1yproprovalaspProaspSer 40
 DB 858 CCCATGATGATGAAGACCAAGATCTTGATCTGATCTGATCTGATCTGATCTGATCTGATCT 917
 QY 41 AspAsnSer1a1e1ytrVal1G1ng1YleuAsnaspSer1a1ThrlaAspApleu1a 60
 DB 918 GACAAACAGTGAATTTATGTCAGAAAGATTAATACAGTGTGACTCTTAATGATCTGGCA 977
 QY 61 AspPhePhe1ySg1ncYsG1yVal1yMetAsn1ySArGThrG1yG1nPromet1le 80
 DB 978 GACTCTTTAAGCAGTGTGGGTGTTTAAGATGAACAAGAACTGGGCAACCATATC 1037
 QY 81 H1e1e1ytrleuAplySg1yThrG1yLysPro1ySg1yAsp1a1ThrValSer1yGlu 100
 DB 1038 CACATCTACCTGGCAAGAAACAGAAAGCCCAAGCGATGACACAGTCTCTATGAA 1097
 QY 101 AspProProThr1a1yVal1a1a1Val1G1uThrPheaspG1yLysaspPheG1n1ySer 120
 DB 1098 GACCCACCACTGCGCAAGAGCTGCCGTGGAAATGTTGATGGAAAGATTTTCAAGGAGAC 1157
 QY 121 Lysleu1ySVal1Ser1eua1aArg1yLysPro1yMetAsnSerMetArg1yG1yLeu 140
 DB 1158 AAATTAAAGTCTCCTTGTCTCGAAGAGCTCCATTAAGCACTTAAGCGGGGTGCTG 1217
 QY 141 ProPro1yG1uG1yArg1yMetProProPro1yLeuArg1yG1yProG1yG1yProG1y 160
 DB 1218 CCACCCCGTGAAGGAGGAGGATGCCACACACCTCCGTGAGGTCCAGAGAGGCCAGGA 1277
 QY 161 G1yProG1yG1yProMetG1yArgMetG1yArg1yG1yYAspArgG1yG1yPhePro 180
 DB 1278 GGTCTTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1337
 QY 181 ProArg1yProArg1ySerArg1yAsnProSerG1yG1yYAsnVal1G1nH1aArg 200
 DB 1338 CCAAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1397
 QY 201 AlaG1yAsp1yTrpG1nCySProAsnProG1yCySg1yAsnG1nAsnPhe1a1TrpArg1y 220
 DB 1398 GCTGAGAGAGTGGAGTGTCCCAATCCGGGTGTGGAAACCAAGAACTTCCGTGAGAGAA 1457

QY 221 GluCyAsnG1nCySelyVal1aPro1ySProG1uG1yPhe1eupProProPheProPro 240
 DB 1458 GACTGACCAAGTGTAAAGGCCCAAGAGCTTCTCCGCGCACCCCTTCCGCC 1517
 QY 241 ProG1yG1yAspArg1yArg1yG1yProG1yG1yMetArg1yG1yArg1yG1yLeu 260
 DB 1518 CCGGAGTGTATGTCGTGGAGAGAGTGGCCCTGTGGCATCGGAGAGAGAGAGTGGCTC 1577
 QY 261 MetAspArg1yG1yProG1yG1yMetPheArg1yG1yArg1yG1yAspArg1yG1y 280
 DB 1578 ATGATGTCGTGTGTCCGAGTGAATGTCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1637
 QY 281 PheArg1yG1yArg1yMetAspArg1yG1yPheG1yG1yArg1yG1yArg1yG1yPro 300
 DB 1638 TTCCGTGTGTGTCCGAGGAGTGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1697
 QY 301 G1yG1yProProG1yProLeuMetG1uG1nMetG1yG1yArgArg1yG1yArg1yG1y 320
 DB 1698 GGGGGGCCCCCTGGACCTTGTATGAAACAGATGGAGAGAAAGAGAGAGAGAGAGGAGGA 1757

RESULT 2

US-09-822-830A-49/c
 Sequence 49, Application US/09822830A
 Patent No. US20020142952A1
 GENERAL INFORMATION:
 APPLICANT: Genetics Institute, Inc.
 APPLICANT: Wong, Gordon G.
 APPLICANT: Clark, Hilary
 APPLICANT: Rechele, Kim
 APPLICANT: Agostino, Michael J.
 APPLICANT: Howes, Steven H.
 APPLICANT: Resnick, Richard J.
 APPLICANT: Gulukota, Kamalakara
 APPLICANT: Graham, James R.
 TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING NOVEL SECRETED PROTEINS
 FILE REFERENCE: GIN 6402
 CURRENT APPLICATION NUMBER: US/09/822,830A
 CURRENT FILING DATE: 2001-03-29
 PRIOR APPLICATION NUMBER: 60/195,604
 PRIOR FILING DATE: 2000-04-06
 NUMBER OF SEQ ID NOS: 631
 SOFTWARE: PatentIn Ver. 2.0
 SEQ ID NO 49
 LENGTH: 2176
 TYPE: DNA
 ORGANISM: Homo sapiens
 US-09-822-830A-49

Alignment Scores:
 Pred. No.: 4.34e-155 Length: 2176
 Score: 1922.00 Matches: 338
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 9 Gaps: 0

US-10-791-017a-2_COPY_319_656 (1-338) x US-09-822-830A-49 (1-2176)

QY 1 G1YGYAARG1YGLYMETG1YSERA1AG1YGLUARG1YGLYPHEANLYSPROGLYGLY 20
 DB 1207 GGAGGAGCGCGTGGAAATGGGAGCGCTGGAGAGAGTGGCTTCATTAAGCCGTGGGA 1148
 QY 21 Prometaspgluclg1yproaspbleuaspleug1yproprovalaspProaspSer 40
 DB 1147 CCCATGATGAAGACCAAGATCTTGAATCTAGGCCCACTTAATCAAGATGAAGACTCT 1088
 QY 41 AspAsnSer1a1e1ytrVal1G1ng1YleuAsnaspSer1a1ThrlaAspApleu1a 60
 DB 1087 GACAAACAGTGAATTTATGTCAGAAAGATTAATACAGTGTGACTCTTAATGATGATCTGCA 1028

QY 61 AspphepheyglnCyvsglyValValLysMetAsnlyPheArgThrGlyGlnPromeC11e 80
DB 1027 GACCTCTTTAAGCAGGTGGGTGTTTAAATGAAAGAACTGGGCAACCATGATC 968
QY 81 HisIIeTyeuAspLysGluThrGlyLysProLysGlyAspAlaThrValSerTYGlu 100
DB 967 CACATCTAAGCTGCAAGAAAGAAAGAAAGCCCAAGCCATGCTGCAAGTCTTATGAA 908
QY 101 AspproProThrAlaLysValAlaValGluTrpPheAspGlyLysAspPheGlnGlySer 120
DB 907 GACCCACCCACCTGCAAGGCTGCGGTGAATGTTTATGAGAAAGATTTTCAAGGAGC 848
QY 121 LysLeuLysValSerLeuAlaArgLysLysProLysMetAsnSerMetArgGlyLysLeu 140
DB 847 AAACCTTAAGTCTCCCTTGTCTCGAAGAACCTTCAATGAAAGATGCGGGGTGCTG 788
QY 141 ProProArgGluGlyArgGlyMetProProProLeuArgGlyLysProGlyLysProGly 160
DB 787 CCAACCCGTGAGGGGCAAGGCAATCCACCACTCCGTGAGGTCCAGAGGCCCCAGGA 728
QY 161 GlyProGlyLysProMetGlyArgMetGlyLysArgGlyLysAspArgGlyLysPhePro 180
DB 727 GGTCTGGGGGAGCCCAATGGGTGCGCATGGAGGCGGTGAGAGATAGAGAGGCTTCCCT 668
QY 181 ProArgGlyLysProArgGlySerArgGlyLysAspProSerGlyLysLysValGlnHisArg 200
DB 667 CCAAGAGGAGCCCGGGGTTCCTCCAGAGGAGCCCTCTGAGAGAGAAAGCTCCAGACCGA 608
QY 201 AlaGlyAspTrpGlnCyvsglyProAspProGlyCyvsglyValGlnHisPheAlaTrpArgThr 220
DB 607 GCTGAGAGCTGCGAGTGTCCCATCCGGGTGTGGAACCAAGATTCCTGCTGAGAGAA 548
QY 221 GluCyvAsnGlnCyvsglyValAspLysProGlnGlyLysPheLeuProProPhePro 240
DB 547 GAGTGCACACCAAGTGAAGGCCCCCAAGCCCTGAAGGCTTCCCTCCGCAACCTTCCGCC 488
QY 241 ProGlyLysAspArgGlyLysArgGlyLysProGlyLysMetArgGlyLysArgGlyLysLeu 260
DB 487 CCGGGGTGTGATGTGTGCAAGAGGTGGCTGTGTGTGCGGGGAGAGAAAGGTGGCTC 428
QY 261 MetAspArgGlyLysLysProGlyLysMetPheArgGlyLysArgGlyLysAspArgGlyLys 280
DB 427 ATGATCTGTGTGTGTCCCGGTGGAATGTTCAGAGGTGGCGGTGTGAGCAAGAGGTGC 368
QY 281 PheArgGlyLysArgGlyLysMetAspArgGlyLysPheGlyLysGlyLysArgGlyLysPro 300
DB 367 TTCGAGGTGTGCGGGGCAATGAGACCAAGGTGGCTTGTGTGAGAGAAAGCAAGGTGGCT 308
QY 301 GlyLysLysProProGlyLysProLeuMetGlnMetGlyLysArgArgGlyLysArgGlyLys 320
DB 307 GGGGGGCCCCCTGTGACCTTTGATGGAACAGATGGAGAGAAAGAGAGGACGCTGGAGGA 248
QY 321 ProGlyLysMetAspLysGlyLysLysArgGlnGlnArgArgAspArgProTYR 338
DB 247 CCTGGAATAATGATTAAGGCGAGCACGTCGAGAGCGCAGAGATCGGCTTAC 194

RESULT 3
US-09-880-107-3769
; Sequence 3769, Application US/09880107
; Patent No. US2002014281A1
; GENERAL INFORMATION:
; APPLICANT: Horne, Darci T.
; APPLICANT: Vockley, Joseph G.
; APPLICANT: Scherf, Uwe
; APPLICANT: Gene Logic, Inc.
; TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
; FILE REFERENCE: 44921-5028-WO
; CURRENT APPLICATION NUMBER: US/09/880,107
; CURRENT FILING DATE: 2001-06-14
; PRIOR APPLICATION NUMBER: US 60/211,379
; PRIOR FILING DATE: 2000-06-14
; PRIOR APPLICATION NUMBER: US 60/237,054

; PRIOR FILING DATE: 2000-10-02
; NUMBER OF SEQ ID NOS: 3950
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3769
; LENGTH: 2390
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURe: Genbank Accession No. US20020142981A1 X66899
; US-09-880-107-3769

Alignment Scores:
Pred. No.: 4,73e-155 Length: 2390
Score: 1922.00 Matches: 338
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 9 Gaps: 0

US-10-791-017a-2_COPY_319_656 (1-338) x US-09-880-107-3769 (1-2390)

QY 1 GlyLysArgGlyLysMetGlySerAlaGlyLysArgGlyLysPheAsnLysProGlyLys 20
DB 998 GAGAGAGCGGTGATGGCAAGGCTGAGAGAGAGGTGCTTCAATAGCTGTGGA 1057
QY 21 ProMetAspGluGlyProAspLeuAspLeuGlyProProValAspProAspGluAspSer 40
DB 1058 CCCATGATTAAGAGACAGATCTTGTATCTAGGCCCTCTGTATCCAGATGAAGACTCT 1117
QY 41 AsppAspSerAlaLeuTrpValGlnGlyLysAsnAspSerValThrLeuAspLeuAla 60
DB 1118 GACACAGTGCATTTATATGTAACAAGATTAATGACAGTGTACTAGATGATCTGGCA 1177
QY 61 AspphepheyglnCyvsglyValValLysMetAsnlyPheArgThrGlyGlnPromeC11e 80
DB 1178 GACTTCTTTAAGCAGGTGGGTGTTTAAATGAAAGAACTGGGCAACCATGATC 1237
QY 81 HisIIeTyeuAspLysGluThrGlyLysProLysGlyLysAspAlaThrValSerTYGlu 100
DB 1238 CACATCTAAGCTGCAAGAAAGAAAGCCCAAGGCGCATGCCAGTGTCTTATGAA 1297
QY 101 AspproProThrAlaLysValAlaValGluTrpPheAspGlyLysAspPheGlnGlySer 120
DB 1298 GACCCACCCACCTGCAAGGCTGCGGTGAATGTTTATGAGAAAGATTTTCAAGGAGC 1357
QY 121 LysLeuLysValSerLeuAlaArgLysLysProProMetAsnSerMetArgGlyLysLeu 140
DB 1358 AAACCTTAAGTCTCCCTTGTCTGGAAGAACCTTCAATGAAAGATGCGGGGTGTGCTG 1417
QY 141 ProProArgGluGlyLysArgGlyLysMetProProProLeuArgGlyLysProGlyLysProGly 160
DB 1418 CCAACCCGTGAGGGGCAAGGCAATGCCACCACTCCGTGAGAGGTCCAGAGGCCCCAGGA 1477
QY 161 GlyProGlyLysProMetGlyLysArgMetGlyLysArgGlyLysAspArgGlyLysPhePro 180
DB 1478 GGTCTGGGGGAGCCCAATGGGTGCAATGGAGGCGGTGTGAGAGAAATGAGAGGCTTCCCT 1537
QY 181 ProArgGlyLysProArgGlyLysSerArgGlyLysAspProSerGlyLysLysValGlnHisArg 200
DB 1538 CCAAGAGAGCCCGGGGTTCCTCCAGAGAACCCCTCTGAGAGAGAAAGCTCCAGACCGA 1597
QY 201 AlaGlyAspTrpGlnCyvsglyProAspProGlyCyvsglyValGlnHisPheAlaTrpArgThr 220
DB 1598 GCTGAGAGCTGCGAGTGTCCCATCCGGGTGTGGAACCAAGAACTTCCGCTGAGAGAA 1657
QY 221 GluCyvAsnGlnCyvsglyValAspLysProGlnGlyLysPheLeuProProPhePro 240
DB 1658 GAGTGCACACCAAGTGAAGGCCCCCAAGGCTGAAAGGCTTCTCCGCAACCTTCCGCC 1717
QY 241 ProGlyLysAspArgGlyLysArgGlyLysProGlyLysMetArgGlyLysArgGlyLysLeu 260
DB 1718 CCGGGGTGTGATGTGTGCAAGAGGTGGCTCTGTGTGATGCGGGGAGAGAAAGGTGGCTC 1777

```
OY 261 MetAspArgGlyValProGlyGlyMetPheArgGlyValArgGlyValAspArgGlyVal 280
DB 1778 ATGATCGTGTGTGTCCGGTGGAAATGTATGAGGTGGCCGTGTGTGAGCAAGAGTGGC 1837
OY 281 PheArgGlyValArgGlyMetAspArgGlyValPheGlyValValArgArgGlyValPro 300
DB 1838 TTCCTGTGTGTGGCGGGGCATGAGCCGAGGTGGCTTTGTGTGAGGAAGAGAGGTGGCCCT 1897
OY 301 GlyValProProGlyProLeuMetGlyGlyMetGlyValArgArgGlyValArgGlyVal 320
DB 1898 GGGGGGGCCCCCTGACCTTTGATGGAACAGATGGAGGAAGAGAGAGAGACCTGTGAGGA 1957
OY 321 ProGlyValMetAspArgGlyValHisArgGlyGlyArgArgArgArgProPro 338
DB 1958 CCGGAAAAATGATTAAGGCGAGCACGTCAGAGGCGCAGAGTCCGCTTAC 2011

RESULT 4
US-09-960-706-1081
; Sequence 1081, Application US/09960706
; Publication No. US20030134280A1
; GENERAL INFORMATION:
; APPLICANT: Munger, William E.
; TITLE OF INVENTION: Identifying Drugs for and Diagnosis of Benign Prostatic Hyperplasia
; FILE REFERENCE: 44921-5029-01US
; CURRENT APPLICATION NUMBER: US/09/960,706
; CURRENT FILING DATE: 2001-09-24
; PRIOR APPLICATION NUMBER: 60/223,323
; PRIOR FILING DATE: 2000-08-07
; PRIOR APPLICATION NUMBER: 09/873,319
; PRIOR FILING DATE: 2001-06-05
; NUMBER OF SEQ ID NOS: 1124
; SOFTWARE: Patent Ver. 2.1
; SEQ ID NO 1081
; LENGTH: 2390
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20030134280A1 X66899
US-09-960-706-1081

Alignment Scores:
Pred. No.: 4,73e-155 Length: 2390
Score: 1922.00 Matches: 338
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 10 Gaps: 0

US-10-791-017a-2_COPY_319_656 (1-338) x US-09-960-706-1081 (1-2390)
OY 1 GlyValArgGlyValMetGlySerAlaGlyValArgGlyValPheAsnValProGlyVal 20
DB 998 GAGAGAGCGCGGTGGAATGGGCGAGCGGTGAGAGCGGAGGTGGCTTCAATTAAGCCGTGGA 1057
OY 21 PrometAspArgGlyValProAspLeuAspLeuGlyProProValAspProAspGlyAspSer 40
DB 1058 CCCATGATGAAGACCAAGATCTTGAATGAGCCCTCTGTAGATCCAGATGAAGACTCT 1117
OY 41 AspaAsnSerAlaIleValGlnGlyLeuAsnAspSerValThrLeuAspAspLeuAla 60
DB 1118 GACACAGGTGCAATTTATGTAACAGATTAATATACAGTGTGACTCTAGATGATCTGGCA 1177
OY 61 AspPhePheValGlnValValValMetAsnValArgThrGlyGlnProMetIle 80
DB 1178 GACTTCTTTAAGCGGTGGGTGTGTTAATGTAACAGAGAACTGGGCAACCATGATC 1237
OY 81 HisIleValArgGlyValValValValValValValValValValValValValValVal 100
DB 1238 CACATCTAAGTGGCAAGAAAGAGAAAGCCCAAGAGCCCAATGCCAAGTGTCTTAAGAA 1297
OY 101 AspProProThrAlaValAlaValAlaValAlaValAlaValAlaValAlaValAlaVal 120
```

```
DB 1298 GACCCACCCACTGCGCAAGGCTCCGATGATGTTGATGGGAAAGATTTTCAGAGGAGC 1357
OY 121 ValLeuValValSerLeuAlaArgValValProProMetAsnSerMetArgGlyValLeu 140
DB 1358 AAACCTTAAGTCTCCCTTGTGTAAGAGCCCTCCATATACAGATGTGGGGTGTCTG 1417
OY 141 ProProArgGlyValArgGlyMetProProProLeuArgGlyValProGlyValProGly 160
DB 1418 CACACCCGTGAGAGGCGAGAGGCATGCCACCACTCTCGTGAAGTCCAGAGGCCCAAGA 1477
OY 161 GlyProGlyValProMetGlyValArgMetGlyValArgGlyValAspArgGlyValPhePro 180
DB 1478 GGTCTGGGGAGCCCATGTGGTGCATGGAGGCCGTGGAGAGATAGAGAGGCTTCCCT 1537
OY 181 ProArgGlyProArgGlySerArgValAsnProSerGlyValGlyValValValValVal 200
DB 1538 CCAAGAGAACCCCGGGGTTCGAGAGGAACCCCTCTGAGAGAGAAACGTCGACACCGA 1597
OY 201 AlaGlyAspTrpGlnCybProAsnProGlyCybGlyAsnGlnAsnPheAlaTrpArgThr 220
DB 1598 GCTGAGACTGCGACAGTGTCCCATCCGGTTGTGAAACCAAGACTTCCGCTGAGAAACA 1657
OY 221 GluCybAsnGlnCybValAlaProValProGlyGlyPheLeuProProPheProPro 240
DB 1658 GAGTGCACACAGTGTAAAGCCCAAGAGCTTCTCCGCGCACCCCTTCCGCCCC 1717
OY 241 ProGlyValAspArgGlyValArgGlyValProGlyValMetArgGlyValArgGlyValLeu 260
DB 1718 CCGGGTGTGATCGTGGCAGAGGTGGCCCTGTGGCATCGGGGAGAGAGAGGTGGCTC 1777
OY 261 MetAspArgGlyValProGlyGlyMetPheArgGlyValArgGlyValValAspArgGlyVal 280
DB 1778 ATGATCGTGTGTGTCCCGGTGGAATGTTCAAGAGTGGCCGTGTGAGACAGAGGTGGC 1837
OY 281 PheArgGlyValArgGlyMetAspArgGlyValPheGlyValValArgArgGlyValPro 300
DB 1838 TTCCTGTGTGTGGCGGGGCATGAGCCGAGGTGGCTTTGTGAGGAAGACAGAGGTGGCCCT 1897
OY 301 GlyValProProGlyProLeuMetGlyGlyMetGlyValArgArgGlyValArgGlyVal 320
DB 1898 GGGGGGGCCCCCTGACCTTTGATGGAACAGATGGAGGAAGAGAGAGAGACCTGTGAGGA 1957
OY 321 ProGlyValMetAspArgGlyValHisArgGlyGlyArgArgArgArgProPro 338
DB 1958 CCGGAAAAATGATTAAGGCGAGCACGTCAGAGGCGCAGAGTCCGCTTAC 2011

RESULT 5
US-09-873-319-717
; Sequence 717, Application US/09873319A
; Publication No. US20030134324A1
; GENERAL INFORMATION:
; APPLICANT: Munger, William E.
; APPLICANT: Kulkarni, Prakash
; APPLICANT: Getzenberg, Robert H.
; APPLICANT: Waga, Iwao
; TITLE OF INVENTION: Identifying Drugs for and Diagnosis of Benign Prostatic
; FILE REFERENCE: 44921-5029-01US
; CURRENT APPLICATION NUMBER: US/09/873,319A
; EARLIER FILING DATE: 2001-06-05
; EARLIER APPLICATION NUMBER: US 60/223,323
; NUMBER OF SEQ ID NOS: 755
; SOFTWARE: Patent Ver. 2.1
; SEQ ID NO 717
; LENGTH: 2390
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20030134324A1 X66899
US-09-873-319-717
```

Alignment Scores:

Pred. No.:	4,73e-155	Length:	2390
Score:	1922.00	Matches:	338
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	10	Gaps:	0

US-10-791-017a-2_COPY_319_656 (1-338) x US-09-873-319-717 (1-2390)

```
QY 1 G1YGLYARGLYGLYMETGLYSERIALGLYLUARGGLYLYPHEANLYSPROGLYGLY 20
DB 998 GGAGGAGCGCGTGAATGGGCAAGGCGTGGAGAGGAGGTGGCTTCATTAAGCCGTGGCA 1057
QY 21 PROMETASPGIUGLYPROASPPLUAPLEUGLYPROBVALASPPOAPSPGLUAPSPER 40
DB 1058 CCCATGGATGAGGACAGATCTTGATCTAGGCCCTCTGTAGATCCAGATGAAGACTCT 1117
QY 41 ASPANSERALIETRYVALINGLYLEUASNPSEVALTHIRLEUASPAAPLEUALA 60
DB 1118 GACAAACGTCGCAATTATGTAACAAGATTAAATGACGTGTGACTTAAATGATGATGGCA 1177
QY 61 ASPPHEHELISGLINCYSGLYVALIYALYMEASNLYSARQTTHRGILYGLINPROMETILE 80
DB 1178 GACTTCTTTAAGCAGTGTGGGTTTGAATGACAGAGAACTGGGCAACCATGATC 1237
QY 81 HISILETRYLEUASPLYSGLIUTHRGILYLYSPROLYSGLYASPLATHVALISERTYRGILU 100
DB 1238 CACATCTAAGCTGGCAAGAGAAACAGAAAGCCCAAGCCATGACAGTGTCTTATGAA 1297
QY 101 ASPPROPTHRALIALYSALAAIYALVALITRPHASPLGLYLYSAPPHINGLYLYSER 120
DB 1298 GACCCCAACCACTGCAGAGGCTGCGTGAATGTGTTATGAGAAAGATTTTCAAGGGAGC 1357
QY 121 LYSLEULYVALISERLEUALAARGLYLYSPROBROMETASNPSEVALTHIRLEUASPAAPLEUALA 140
DB 1358 AAACCTTAAGTCTCCCTGCTCGGAGAGAACCTTCATTAAGCACTATGCGGGGTGCTG 1417
QY 141 PROPROARGIUGLYARGLYMETPROPROBROMETASPLGLYLYSPROGLYLYPROGLY 160
DB 1418 CCAACCCGAGGAGCAAGGAGCATGCAACCACTCGTGAAGGTCCAGAGAGGCCCAAGGA 1477
QY 161 GLYPROGLYGLYLYPROMETGLYARGMETGLYLYARGGLYLYASPARGLYGLYLYPHEPRO 180
DB 1478 GGTCTGGGGGAGCCCATGGGTGCGATGCGAGGCGGTGAGAGAGATTAAGAGAGGCTTCCCT 1537
QY 181 PROARGIUGLYPROARGIUGLYSERARGLYLYASNPROSEGLYGLYLYLYASNVALGINHISARG 200
DB 1538 CCAAGAGGAGCCCGGGGTTCCCAAGGAGAACCTCTTGAGAGAGAGAACGTCACACCGCA 1597
QY 201 ALIAGLYASPTIRPOLINCYSPROASNPROGLYLYSGLYLYANGLYLYNANPHEALATIRPARGTHR 220
DB 1598 GCTGGAGACTGGAGGTGTCCCAATCCGGGTTGTGAGAACCAAGAACTTCGCTGGAGAA 1657
QY 221 GLUCYASANGLYLYLYSALAPROLYSPROGLYLYLYPHELEUPROPROPHETPRO 240
DB 1658 GAGTGCACACGATGAAGGCCCAAMAGCTGAAAGGCTTCTCCGCGCAACCTTCCGCCCC 1717
QY 241 PROGLYGLYASPARGLYARGGLYGLYLYPROGLYGLYLYMETARGLYGLYLYARGGLYGLYLYLEU 260
DB 1718 CCGGGGTGTATGTGGCAAGGTGGCCCTGTGTGGCATGCGGGAGAGAGAGGTGGCTTC 1777
QY 261 METASPARGLYGLYLYPROGLYGLYLYMETPHEARGGLYGLYLYARGGLYGLYLYASPARGLYGLYLY 280
DB 1778 ATGAGATGTGTGGTCCCGGAGAAATGTTACAGAGGTGGCCGTGTGTGAGACAGAGGTGGC 1837
QY 281 PHEARGGLYGLYLYARGGLYLYMETASPARGLYGLYLYPHEGLYGLYLYLYARGGLYGLYLYPRO 300
DB 1838 TTCCTGTGTGGTGGCGGAGCATGAGACGAGGTGCTTTGTGTGAGAGAGAACAGAGGTGGCCCT 1897
QY 301 GLYLYPROBROGLYLYPROLEUMETGLYLYNMEGLYGLYLYARGARGGLYGLYLYARGGLYGLYLY 320
DB 1898 GGGGGGGCCCTTGACCTTTGATGGAACAATGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1957
```

```
QY 321 PROGLYLYMEASPLYSGLYGLYLIARGGLYLYLYARGARGASPARGLYLY 338
DB 1958 CCTGGAAAAATGATTAAGCGAGACCGTCAAGAGCGCAGAGATCGGCCCTTAC 2011
```

RESULT 6

```
US-10-791-017a-1
; Sequence 1, Application US/10791017A
; Publication No. US20040197827A1
; GENERAL INFORMATION:
; APPLICANT: JENAPHARM GmbH & Co. KG
; TITLE OF INVENTION: Methods for Determining Hormonal Effects of Substances
; FILE REFERENCE: Pat 3684/11
; CURRENT APPLICATION NUMBER: US/10/791, 017A
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 1
; LENGTH: 2390
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURES:
; NAME/KEY: CDS
; LOCATION: (44)..(2011)
; OTHER INFORMATION: EMS
US-10-791-017a-1
```

Alignment Scores:

Pred. No.:	4,73e-155	Length:	2390
Score:	1922.00	Matches:	338
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	18	Gaps:	0

US-10-791-017a-2_COPY_319_656 (1-338) x US-10-791-017a-1 (1-2390)

```
QY 1 G1YGLYARGLYGLYMETGLYSERIALGLYLUARGGLYLYPHEANLYSPROGLYGLY 20
DB 998 GGAGGAGCGCGTGAATGGGCAAGGCGTGGAGAGGAGGTGGCTTCATTAAGCCGTGGCA 1057
QY 21 PROMETASPGIUGLYPROASPPLUAPLEUGLYPROBVALASPPOAPSPGLUAPSPER 40
DB 1058 CCCATGGATGAGGACAGATCTTGATCTAGGCCCTGTAGATCCAGATGAAGACTCT 1117
QY 41 ASPANSERALIETRYVALINGLYLEUASNPSEVALTHIRLEUASPAAPLEUALA 60
DB 1118 GACAAACGTCGCAATTATGTAACAAGATTAAATGACGTGTGACTTAAATGATGATGGCA 1177
QY 61 ASPPHEHELISGLINCYSGLYVALIYALYMEASNLYSARQTTHRGILYGLINPROMETILE 80
DB 1178 GACTTCTTTAAGCAGTGTGGGTTTGAATGACAGAGAACTGGGCAACCATGATC 1237
QY 81 HISILETRYLEUASPLYSGLIUTHRGILYLYSPROLYSGLYASPLATHVALISERTYRGILU 100
DB 1238 CACATCTAAGCTGGCAAGAGAAACAGAAAGCCCAAGCCATGACAGTGTCTTATGAA 1297
QY 101 ASPPROPTHRALIALYSALAAIYALVALITRPHASPLGLYLYSAPPHINGLYLYSER 120
DB 1298 GACCCCAACCACTGCAGAGGCTGCGTGAATGTGTTATGAGAAAGATTTTCAAGGAGC 1357
QY 121 LYSLEULYVALISERLEUALAARGLYLYSPROBROMETASNPSEVALTHIRLEUASPAAPLEUALA 140
DB 1358 AAACCTTAAGTCTCCCTGCTCGGAGAGAACCTTCATTAAGCACTATGCGGGGTGCTG 1417
QY 141 PROPROARGIUGLYARGLYMETPROPROBROMETASPLGLYLYSPROGLYLYPROGLY 160
DB 1418 CCAACCCGAGGAGCAAGGAGCATGCAACCACTCGTGAAGGTCCAGAGAGGCCCAAGGA 1477
QY 161 GLYPROGLYGLYLYPROMETGLYARGMETGLYLYARGGLYLYLYASPARGLYGLYLYPHEPRO 180
DB 1478 GGTCTGGGGGAGCCCATGGGTGCGATGCGAGGCGGTGAGAGAGATTAAGAGAGGCTTCCCT 1537
```

QY 181 ProArgGlyProArgGlySerArgGlyAsnProSerGlyGlyGlyAsnValGlnHisArg 200
DB 1538 CCAAGAGACCCCGGGGTTCCCAAGGAGAACCTTCGAGAGAGAACGTCACGACCGA 1597
QY 201 AlaGlyAspTrpGlnCysProAsnProGlyCysGlyAsnGlnAsnPheAlaTrpArgThr 220
DB 1598 GCTGGAGACTGGCAGTGTCCCAATCCGGGTTGTGGAACAGAACTTCGCTGGAGACA 1657
QY 221 GluCyAsnGlnCysGlyValAlaProLysProGluGlyPheLeuProProPheProPro 240
DB 1658 GAGTGCACACAGGTAAAGGCCCAAAAGGCTTCCTCCGACACCTTCGCGCC 1717
QY 241 ProGlyValAspArgGlyArgGlyGlyProGlyGlyMetArgGlyArgGlyGlyLeu 260
DB 1718 CCGGCTGTGATCTGTGCAAGGTGCGCTGTGTGCAATGCGGGAGAGAGGTGCTC 1777
QY 261 MetAspArgGlyGlyProGlyGlyMetPheArgGlyGlyValArgGlyGlyAspArgGly 280
DB 1778 ATGGATCGTGTGTGTCCTGGTGAATGTTCAAGAGTGGCCGTGTGAGACAGAGGTGGC 1837
QY 281 PheArgGlyGlyArgGlyMetAspArgGlyGlyPheGlyGlyGlyValArgGlyGlyPro 300
DB 1838 TTCGTGTGTGTCGGGGCAGTGAAGCCGAGGTGTGTTGTGTGAAGAAACAGGTGTGGCT 1897
QY 301 GlyGlyProProGlyProLeuMetGluGlnMetGlyGlyValArgArgGlyGlyArgGly 320
DB 1898 GGGGGGCCCCCTTGACCTTTGATGAGAACATGTGGAGAGAAAGAGAGAGAGAGAGAG 1957
QY 321 ProGlyValMetAspArgGlyGlyGlnHisArgGlnGluArgArgArgArgProGly 338
DB 1958 CTGGAGAAATGATTAAGCGAGACCGTCAGAGCGAGAGATCGGCTTAC 2011

RESULT 7

US-09-822-830A-410/C
; Sequence 410, Application US/09822830A
; Patent No. US20020142952A1
; GENERAL INFORMATION:
; APPLICANT: Genetics Institute, Inc.
; APPLICANT: Mong, Gordon G.
; APPLICANT: Clark, Hilary
; APPLICANT: Fechtel, Kim
; APPLICANT: Agostino, Michael J.
; APPLICANT: Howes, Steven H.
; APPLICANT: Resnick, Richard J.
; APPLICANT: Gulukota, Kamalakari
; APPLICANT: Graham, James R.
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING NOVEL SECRETED PROTEINS
; FILE REFERENCE: GIN 6402
; CURRENT APPLICATION NUMBER: US/09/822,830A
; CURRENT FILING DATE: 2001-03-29
; PRIOR APPLICATION NUMBER: 60/195,604
; PRIOR FILING DATE: 2000-04-06
; NUMBER OF SEQ ID NOS: 631
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 410
; LENGTH: 2273
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-822-830A-410

Alignment Scores:

Prod. No.:	1,13e-142	Length:	2273
Score:	1777.00	Matches:	319
Percent Similarity:	94.38%	Conservative:	0
Best Local Similarity:	94.38%	Mismatches:	1
Query Match:	92.46%	Indels:	18
DB:	9	Gaps:	2

US-10-791-017a-2_COPY_319_656 (1-338) x US-09-822-830A-410 (1-2273)

QY 1 GlyValArgGlyGlyMetGlySerAlaGlyGlnArgGlyGlyPheAsnLysProGlyGly 20
DB 1312 GGAAGAGCGCGTGAATGGGCG---GCTGAGAGAGGAGGTGGCTTCATTAAGCCGTGTGA 1256

QY 21 ProMetAspGluGlyProAspLeuAspLeuGlyProProValAspProAspGluAspSer 40
DB 1255 CCCATGATGAAGAACCAATCTTGATCTAGGCCACCTGTAATCAAGTAAAGACTCT 1196
QY 41 AspAsnSerAlaIleTyrValGlnGlyLeuAsnAspSerValThrLeuAspLeuAla 60
DB 1195 GACAAAGAGCAATTTATGTACAGAGATTAAATGACAGTGTACTCTAGATGATCTGGCA 1136
QY 61 AspPhePheLeuGlnCysGlyValValLysMetAsnLysArgThrGlyGlnProMetIle 80
DB 1135 GACTCTTTAAGCAGGTGTGGGTGTGTTAAGATGAACAGAGACTGGCAACCAATGATC 1076
QY 81 HisIleTyrLeuAspLeuThrGlyLysProLysGlyAspAlaThrValSerTyrGlu 100
DB 1075 CACATCTACCTGGACAGAAACAGAAAGCCCAAGGCGATGCCACAGTCTCTATGA 1016
QY 101 AspProProThrAlaLysAlaAlaValGluTrpPheAspGlyLysAspPheGlnGlySer 120
DB 1015 GACCCACCACTGCGCAAGGCTGCGTGAATGTTGTTGGAAAGATTTTCAAGGAGAC 956
QY 121 LysLeuLysValSerLeuAlaArgLysLysProProMetAsnSerMetArgGlyGlyLeu 140
DB 955 AAACCTTAAGTCTCCCTTCTCGGAAGAGCTCCATATGACATATGCGGGGTGTCTG 896
QY 141 ProProArgGluGlyArgGlyMetProProProLeuArgGlyGlyProGlyGlyProGly 160
DB 895 CCAACCCCGTGAAGGACAGAGCAGTGCACCACTCGTGAAGGTCCAGAGGCCCAAGGA 836
QY 161 GlyProGlyGlyProMetGlyArgMetGlyGlyArgGlyGlyValAspArgGlyGlyPhePro 180
DB 835 GGTCTGGGGGACCCATGGGTGCGCATGGAGGCGGTGAGAGATGAGAGGCTTCCCT 776
QY 181 ProArgGlyProArgGlySerArgGlyAsnProSerGlyGlyGlyAsnValGlnHisArg 200
DB 775 CCAAGAGACCCCGGGGTTCCCAAGGAAACCTCTGAGAGAGAAACGTCAGACCGA 716
QY 201 AlaGlyAspTrpGlnCysProAsnProGlyCysGlyAsnGlnAsnPheAlaTrpArgThr 220
DB 715 GCTGGAGACTGCGAGTGTCCCAATCCGGGTTGTGGAACAGAACTTCGCTGAGAGACA 656
QY 221 GluCyAsnGlnCysGlyValAlaProLysProGluGlyPheLeuProProPheProPro 240
DB 655 GAGTGCACCAAGT----- 641
QY 241 ProGlyValAspArgGlyArgGlyGlyProGlyGlyMetArgGlyGlyValArgGlyGlyLeu 260
DB 640 -----GTTGATGTGTGAGAGGTGCTCTGTGTGATGCGGGAGAGAGAGGTGGCTC 587
QY 261 MetAspArgGlyGlyProGlyGlyMetPheArgGlyGlyValArgGlyGlyAspArgGlyGly 280
DB 586 ATGGATCGTGTGTGTCCTGGTGAATGTTCAAGAGTGGCCGTGTGAGACAGAGGTGGC 527
QY 281 PheArgGlyGlyArgGlyMetAspArgGlyGlyPheGlyGlyGlyValArgArgGlyGlyPro 300
DB 526 TTCGTGTGTGCGGGGCAATGACAGAGGTGCTTTGTGTGAGAGAGACAGAGGTGGCT 467
QY 301 GlyGlyProProGlyProLeuMetGluGlnMetGlyGlyValArgArgGlyGlyValArgGly 320
DB 466 GGGGGGCCCCCTTGACCTTTGATGAGAACAGATGGAGAGAAAGAGAGAGAGAGAGAGAG 407
QY 321 ProGlyValMetAspArgGlyGlyGlnHisArgGlnGluArgArgArgArgProGly 338
DB 406 CTGGAGAAATGATTAAGCGAGACCGTCAGAGCGAGAGATCGGCTTAC 353

RESULT 8

US-10-198-846-9847
; Sequence 9847, Application US/10198846
; Publication No. US2003009974A1
; GENERAL INFORMATION:
; APPLICANT: Lillie, James
; APPLICANT: Xu, Yongyao
; APPLICANT: Wang, Youzhen

	/	APPLICANT:	Stetlemann, Kathleen	
	/	TITLE OF INVENTION:	NOVEL GENES, COMPOSITIONS, KITS, AND METHODS	
	/	TITLE OF INVENTION:	FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND...	
	/	TITLE OF INVENTION:	THERAPY OF BREAST CANCER	
	/	FILE REFERENCE:	MRI-049	
	/	CURRENT APPLICATION NUMBER:	US/10/198,846	
	/	CURRENT FILING DATE:	2002-07-18	
	/	PRIOR APPLICATION NUMBER:	60/306,220	
	/	PRIOR FILING DATE:	2001-07-18	
	/	NUMBER OF SEQ ID NOS:	14084	
	/	SOFTWARE:	FastSeq for Windows Version 4.0	
	/	SEQ ID NO	9847	
	/	LENGTH:	2239	
	/	TYPE:	DNA	
	/	ORGANISM:	Homo sapiens	
	/		US-10-198-846-9847	
		Alignment Scores:		
		Pred. No.:	1,05e-91	Length: 2239
		Score:	1181.00	Matches: 222
		Percent Similarity:	66.07%	Conservative: 0
		Best Local Similarity:	66.07%	Mismatches: 1
		Query Match:	61.45%	Indels: 114
		Gaps:	14	Gaps: 1
		US-10-791-017A-2_COPY_319_656 (1-338) x US-10-198-846-9847 (1-2299)		
Oy		20	GlyPProMetApBgGluGIyLPProAspleuAspleuGIyLPProProValaspProAspleGIuASP	39
Db		1072	GGAACCAAGATGGTAAAGGAGAACAGATCTTAATCTTAAGGCCACCCTGTGATCCAGATGAACAAC	1131
Oy		40	SerAspaSenSerAlaIleTyxValGIingLyLeuaAnaSpSerValThrLeuaSpAspleu	59
Db		1132	TCTGCACACAGTGCAATTATGTACTACAGATTTAAATGACAGGTGACTCTAGATGATCTG	1191
Oy		60	AlaAspPhePhelyegInCyegIyValValIyMwtaenLyBArtghrGIyLnProMet	79
Db		1192	GCAGACTTCCTTAAGCACGTGTGGGGTGTAAAGATGAACAAGAGAATCGGCCAACCATYG	1251
Oy		80	IleHsiIleTyxTrleuAspleuGIuThrGIySPProLySGIyASpAlaIthrValSerTyx	99
Db		1232	ATCCCATCTACCTGAGACAGGAACAGAAAGCCCAAGGGAGATGCCACAGTGTCTAT	1311
Oy		100	GIuAspProProThralAlaynAlaIalValGIuTPheAspGIyLYAspPheGIingLy	119
Db		1312	GAAAGCCACCACTGCCACAGGCTGCCGTGGAAATGTTGATGGAAAAATTTCAAGG	1371
Oy		120	SerLybLeuLyValSerLeuAlaArgLybLySPProMetAsnSerMetArGIyGIy	139
Db		1372	AGCAAACCTTAAGTCTCCCTGTCTCTGGAAAGCCTCCAATGAACAGTATCGGGGTGT	1431
Oy		140	LeuProProArGIuGIyARgGLyMePProProPoleuArGIyGIyProGIyGIyPro	159
Db		1432	CTGCCACCCCCTGAAGGGAGAGGCAATGCCACCACTCCGTGGAGGTCCAGAGGCCCA	1491
Oy		160	GIyGIyProGIyGIyProMetGIyARgMeGIyGIyARgGIyGIyASpArGIyGIyPhe	179
Db		1492	GGAAGTCTCTGGGGGAGCCATGGGTGTGCATGGGAGGCCGTGGAGAGATATAGAGAGCTTC	1551
Oy		180	ProProArGIyProArGIySerArGIyAsnPProSerGIyGIyGIyAsnValGIinhis	199
Db		1552	CCTCCMAAGAGCCCCGGGGGTTCCTCGAAGGAACCCCTCTGGAGGGAGAAAGTCCAGAC	1611
Oy		200	ArgAlaGIyAspTrpGIyCysProAsnProGIy-----	210
Db		1612	CGAGCTGAGACTGGCAGTGTCCCAATCCGTA-TGTACTTGTCTTGGCAAAATTGATACC	1670
Oy		210	-----	210
Db		1671	TACGAGTAGAGCACCTTCCCTCACCCCATCCCACTGAAGAGATGTGCTGTCTA	1730
Oy		210	-----	210

Db		1731	GAGAGAAGCAGTANTGATGACCCGTGATGGCTGGTTAGGAGACACTAGTCAGCATTACTGGAC	1790
Qy	/	210	-----	210
Db		1791	GCTTCAGAGCCTTGCGTAGAAGATTGTGAACCTGCTCTGTGGGTCAATGCTCCTGAGGC	1856
Qy		210	-----	210
Db		1851	TGTGCCCTAAAGCATGGGTGTACATAGATCTCTTGATAGTGTGTACTCTTACA	1910
Qy		210	-----	210
Db		1911	CACGACCTTCTGTTATCTTCCTCCTAGTTCATATGGTATTTCTGCTGTGATGTAT	1970
Qy		211	-----CYGLYASRGLNAsnPhenAlaTrpArgThrIleUCysAsnGlnCysLeu	226
Db		1971	GTATGACAGGGGTGTGGGAAACCAAACTTCGCTCGGAGAACAGAGTGTCAACAGTGTAAAG	2030
Qy		227	ALAProLysPProGluGlyPheLeuProProProPheProProGly	242
Db		2031	GCCCCAAGCCTGAAGGCTTCTCTCCCGCACCTTTCCGCCCCGGGT	2078
RESULT 9				
US-09-918-995-27690				
; Sequence 27690, Application US/09918995				
; Publication No., US20030073623A1				
GENERAL INFORMATION:				
APPLICANT: Hyseq, Inc.				
TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED				
TITLE OF INVENTION: FROM VARIOUS CDNA LIBRARIES				
FILE REFERENCE: 20411-756				
CURRENT APPLICATION NUMBER: US/09/918, 995				
PRIOR FILING DATE: 1999-01-20				
PRIOR APPLICATION NUMBER: US/09/235, 076				
NUMBER OF SEQ ID NOS: 38054				
SOFTWARE: FastSeq for Windows Version 3.0				
SEQ ID NO 27690				
LENGTH: 550				
TYPE: DNA				
ORGANISM: Homo sapiens				
FEATURE:				
NAME/KEY: misc feature				
LOCATION: (1) ..(550)				
OTHER INFORMATION: n = A,T,C or G				
US-09-918-995-27690				
Alignment Scores:				
Pred. No.:	1,86e-73	Length:	550	
Score:	961.00	Matches:	166	
Percent Similarity:	97.65%	Conservative:	0	
Best Local Similarity:	97.65%	Mismatches:	4	
Query Match:	50.00%	Indels:	0	
DB:	10	Gaps:	0	
US-10-791-017A-2_COPY_319_656 (1-338) x US-09-918-995-27690 (1-550)				
Qy		119	GLYSerLYSerLeuLYValSerLeuAlaArgLYSyrProPromtAsnSerMetArgLY	138
Db		40	GGAGAGCAACTTANAGTCTCCCTTGCTGGAGAGGCTCCAAAGAACAGATGGGTGT	99
Qy		139	GLYLeuProPArGArgLUglYArGlyMeCPProProLeuArGlyGLYProGLYGLY	158
Db		100	GGTCTGCACACCCTGAGGGGAGCGCATGCGATCGAGAGCCGTGAGAGGTCCAGAGGC	159
Qy		159	ProGLYglYProGLYglYPrometGLYArgmetGLYlYArgGLYlYAspArGLYGLY	178
Db		160	CCAGAGAGTCTCTGGGGGAGCCCATGGGTGCGATGAGAGCCGTGAGAGATAGAGAGGC	219
Qy		179	PheProPArGArgLYProArGArgLYSerArgGLYAsnProSerGLYGLYlYAsnValGln	198
Db		220	TTCCCTCCMAAGGAGCCCGGGGATTTCCCGAGGAGAACCCCTCTTGAGAGAGGAAACGTCCAG	279

```

QY      199  HSAAGALAGLYASPTROGLNCPProAnPProGLVYSGLYAsnGlnAmpnPhealATP 218
Db      280  CACCAAGCTGGAGACTGGCAGGTGTCCCATTCGGGTGTGGAAACAAGAATTGCCCTGG 339
QY      219  ArgThrgLynCyAsnGlnCyAlaValAProLyPProGLUGlyPheLeuProProPhe 238
Db      340  AGAACAGAGTGCACACAGGTGAAGGCCCAAAGACCAGGAGCTTCCTCCCGCACCTTT 399
QY      229  ProProProProGLYAlaSPArgrGLyArGrGLYmeLArgGLYAlaArgGLY 258
Db      400  CGCCCCCGGGGGTGATGTGTGGCAAGGTGGCCCTGTGGCAATGGAGAGAGGT 459
QY      259  GYLmeuMeAspArgLYglYPProGLYmeLPheArgGLYglYargGLYlaAParG 278
Db      460  GGCTTCATGATCGTGTGTGTCCCGTGGAAATTTATAGTGTGCCGTGTGAGACAGA 519
QY      279  GLYglYPheArgGLYglYargGLYmeLAp 288
Db      520  GGTTGGCTTCGTGTGGCCGGGCGCATGAGAC 549

RESULT 10
US-10-755-889-649
; Sequence 649, Application US/10755889
; Publication No. US20040171823A1
GENERAL INFORMATION:
APPLICANT: Bristol-Myers Squibb Company
TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES ASSOCIATED WITH THE NF-KB
FILE OF INVENTION: PATHWAY
FILE REFERENCE: D0284 NP
CURRENT APPLICATION NUMBER: US/10/755, 889
CURRENT FILING DATE: 2004-01-13
PRIOR APPLICATION NUMBER: U.S. 60/440, 068
PRIOR FILING DATE: 2003-01-14
PRIOR APPLICATION NUMBER: U.S. 60/469, 757
PRIOR FILING DATE: 2003-05-12
NUMBER OF SEQ ID NOS: 823
SOFTWARE: PatentIn version 3.2
SEQ ID NO 649
LENGTH: 1822
TYPE: DNA
ORGANISM: Homo sapiens
US-10-755-889-649

Alignment Scores:
Pred. No.:          6,92e-66           Length:          1822
Score:              878.00            Matches:         182
Percent Similarity: 63.37%             Conservative:    36
Best Local Similarity: 52.91%           Mismatches:     54
Query Match:        45.68%             Indels:         72
DB:                 18                  Gaps:           13

US-10-791-017A-2_COPY_319_656 (1-338) x US-10-755-889-649 (1-1822)
QY      1  GIgLYARGlYGLYmEtGlSeraLaGIguARgLIgLYpHeAnLYPProGLYglY 20
Db      814  GGAGGAGAGGTGGCATGGCGGAGAGT---GACCGTGTGGCTTCAATAAATTTGGTGGC 870
QY      21  PromEAspGLUGlyPProAspleuAPleuGLYProProVALAsPProAspGLUAspSer 40
Db      871  CCTCGGAGACCAAGATCAGCTCATGAC-----TCCGACACAGATTAATTC 915
QY      41  AspaSeSaLaIleTYrVaLIgInGLYLeuaNsPSeSaValThrlEuAsPaLleuLa 60
Db      916  GACACACAAACATCTTTGTGCAAGCCTGGTGAAGATTTCAATTAGTCTGTGGCT 975
QY      61  AspPhePhelYgInCYsgLYvalValLYmeLAsnLYsaRThrgLInPrOmELile 80
Db      976  GATTACTTCAAGCAGATTGGTATTATTAAAGACAAACAGAAACGGGACAGCCCATGATT 1035
QY      81  HIsLleTYrLeuaSpLYseGuThrgLYyrPProLYgLYAsPaLaThrVALserTYrGU 100
Db      1036  AATTGTATACAGACAGCGGGAACCTGGCAAGCTGAAGGAGAGGCAACGCGTCTTTGAT 1095

```

Oy AspProcthrAlaValAlaValAlaValGluTrpPheAspGlyLysAspPheGlnGlySer 120
 101 AspProcthrAlaValAlaValAlaValGluTrpPheAspGlyLysAspPheGlnGlySer 120
 1096 GACCACCTTCAGCTTAAGCAGCATGACTGCTTATGCTTAAGTAAAGAAATTCGCGGAAT 1155
 Oy LysLeuLysValSerLeuAlaArgLysLysProProMetCysSerMetArgGlyLysLeu 140
 121 LysLeuLysValSerLeuAlaArgLysLysProProMetCysSerMetArgGlyLysLeu 140
 Db CCTATCAAGGCTCATTTGCTACTACGCGCGGACAGACTTTAAT-----CGGGGTGTGTGC 1209
 Oy ProProArgGluGlyValArgGlyMetProProProLeuArgGlyLysProGlyLysProGly 160
 141 ProProArgGluGlyValArgGlyMetProProProLeuArgGlyLysProGlyLysProGly 160
 Db AARGGTGTGAGGCGCCAGGG-----CGAGAGAGAGCCATGAGCGGTGGA 1254
 Oy GlyProGlyLysProMetGlyArgMetGlyLysValArgGlyLysAspArgGlyLysPhePro 180
 161 GlyProGlyLysProMetGlyArgMetGlyLysValArgGlyLysAspArgGlyLysPhePro 180
 Db GGCTATGAGGTGTGT-----GGCAGTGTGTGTGTGTGTGCGCCAGAGAGATTTTCC 1302
 Oy ProArgGlyProArgGlySerArgGlyAsnProSerGlyLysValGlnHisArg 200
 181 ProArgGlyProArgGlySerArgGlyAsnProSerGlyLysValGlnHisArg 200
 Db AGTGAAGGT-----GATGCGGTGAGAGACAGACGGA 1335
 Oy AlaGlyAspTrpGlnCysProAsnProGlyLysGlnLysGlnAsnPheAlaTrpArgGln 220
 201 AlaGlyAspTrpGlnCysProAsnProGlyLysGlnLysGlnAsnPheAlaTrpArgGln 220
 Db GCTGTGATCAGGAAGTGTCTTAATCCACCTGTGAGAATATGAATCTTCTTGAAGAGAT 1395
 Oy GluCysAsnGlnCysLysValProLysProGluGlyPheLeuProProPheProPro 240
 221 GluCysAsnGlnCysLysValProLysProGluGlyPheLeuProProPheProPro 240
 Db GAATGCAACCATGTTAAGCCCTTAACCAAGATGAC----- 1431
 Oy ProGlyLysAspArgGlyArgGlyLysProGlyLysMetArg-----GlyValArgGlyLys 259
 241 ProGlyLysAspArgGlyArgGlyLysProGlyLysMetArg-----GlyValArgGlyLys 259
 Db CCAAGG-----GGGGACAGGTGTCTCTCATGTGGGGGTAACTAACGGG 1476
 Oy LeuMetAspArgGlyLysProGlyLysMetCysArgGly-----GlyValArgGlyLys 276
 260 LeuMetAspArgGlyLysProGlyLysMetCysArgGly-----GlyValArgGlyLys 276
 Db GATGATCGTCTGTGTGACAGAGAGGCTATGATGAGGCGGCTACCGGGGCGCGCGG 1536
 Oy AsparGlyLysPheArgGlyLysValArg-----GlyMetAspArgGlyLysPheGlyLysGly 295
 277 AsparGlyLysPheArgGlyLysValArg-----GlyMetAspArgGlyLysPheGlyLysGly 295
 Db GACCGTGAAGGCTTCCAGGGGGGCGGGGTGTGGGACAGAGGTGCTTT----- 1587
 Oy ArgArgGlyLysProGlyLysProProGlyProLeuMetGluGlnMetGlyLysArg 315
 296 ArgArgGlyLysProGlyLysProProGlyProLeuMetGluGlnMetGlyLysArg 315
 Db ----- 1587
 Oy GlyLysValArgGlyLysProGlyLysMetCysP-----LysGlyLysHisValArgGlnLysArg 334
 316 GlyLysValArgGlyLysProGlyLysMetCysP-----LysGlyLysHisValArgGlnLysArg 334
 Db -----GGCCTGTGCAATGATTCAGGGGTGACACAGACAGATGACAGG 1635
 Oy 335 AsparProGly 338
 335 AsparProGly 338
 Db 1636 GAGAGCGCGTAT 1647
 1636 GAGAGCGCGTAT 1647
 RESULT 11
 US-10-439-703-58
 ; Sequence 58, Application US/10439703
 ; Publication No. US20040018527A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Chang, Jenny
 ; APPLICANT: O'Connell, Peter
 ; TITLE OF INVENTION: Differential Patterns of Gene Expression that Predict for Docetaxel
 ; TITLE OF INVENTION: Chemoresistivity and Chemoresistance
 ; FILE REFERENCE: HO-P2482US1/10205813
 ; CURRENT APPLICATION NUMBER: US/10/439, 703
 ; PRIOR FILING DATE: 2003-05-16
 ; PRIOR FILING DATE: 2002-05-17
 ; NUMBER OF SEQ ID NOS: 91
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 58
 ; LENGTH: 1824
 ; TYPE: DNA
 ; ORGANISM: Human
 US-10-439-703-58

Alignment Scores:

Pred. No.: 6,93e-66 Length: 1824
 Score: 878.00 Matches: 182
 Percent Similarity: 63.37% Conservative: 36
 Best Local Similarity: 52.91% Mismatches: 54
 Query Match: 45.68% Indels: 72
 DB: 17 Gaps: 13

US-10-791-017a-2_COPY_319_656 (1-338) x US-10-439-703-58 (1-1824)

QY 1 G1G1YARGLYGLYMETGLYSERIALGLYUARGLYGLYPHEANLYSPROGLYGLY 20
 DB 823 GAGGAGAGAGTGGCATGGCCGGAAGT---GACCGTGTGGCTTCATTAATTTGGTGGC 879
 QY 21 ProMetAspGluGlyProAspLeuAspLeuGlyProProValAspProAspGluAspSer 40
 DB 880 CCTCGGAGCCAGGATGATCAGTCATGAC-----TCCGAACAGATTAATTTCA 924
 QY 41 AspAspSerAlaIleTyValGlnGlyLeuAspAspSerValThrLeuAspAspLeuAla 60
 DB 925 GACAAACAACACCATCTTGTGCAAGGCTGGTGAATGTTCATTAATTTGATGCTGCTGCT 984
 QY 61 AspPhePheLeuGlnCysGlyValAlaIleYmeCAsnLysArgThrGlyGlnProMetIle 80
 DB 985 GATTACTTCAAGCAGATGGTATTTATTAACAACAACAGAAACGGACAGCCCATGATTT 1044
 QY 81 HisIleTyLeuAspLysGlnThrGlyLysProLysGlyAspAlaThrValSerTyGlu 100
 DB 1045 AATTGTGACACAGACAGGAACTGGCAGCTGAAGGAGAGGCAACGGCTCTTTGAT 1104
 QY 101 AspProProThrAlaLysAlaAlaValAlaGluTrpPheAspGlyLysAspPheGlnGlySer 120
 DB 1105 GACCCACCTTCACCTTAAGCAGCTATTGACTGTTGATGATAAGAAATTCCTCCGGAAT 1164
 QY 121 LysLeuLysValSerLeuAlaArgLysLysProLysProLysSerMetCArgGlyLysLeu 140
 DB 1165 CCTATCAAGTCTCTATTGCTACTCGCCGGGACAGACTTTAAT---CGGGGTGGTGGC 1218
 QY 141 ProProArgGluGlyArgGlyMetProProProLeuArgGlyGlyProGlyGlyProGly 160
 DB 1219 AATGTCGTGGAGGCCGAGGG-----CGAGAGAGGACCATGGGCGCGGCA 1263
 QY 161 GlyProGlyGlyProMetGlyArgMetGlyLysArgGlyLysAspArgGlyLysPhePro 180
 DB 1264 GGGTATGAGAGTGTGT---GGCAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1311
 QY 181 ProArgGlyProArgLysSerArgLysAsnProSerGlyGlyLysAsnValGlnHisArg 200
 DB 1312 AGTGGAGGT-----GGTGGCGGTGGAGGACAGCAGCGCA 1344
 QY 201 AlaGlyAspTrpGlnCysProAsnProGlyCysGlyAsnGlnAsnPheAlaTrpArgThr 220
 DB 1345 GCTGTGAGCTGAGAGTGTCTTAATCCACTGTGAGAAATATGAATCTTGTGAGGAAT 1404
 QY 221 GluCysAsnGlnCysLysAlaLysProLysProGluGlyPheLeuProProPheProPro 240
 DB 1405 GAATGCAACAGAGTAAAGCCCTTAACCAAGATGCG-----1440
 QY 241 ProGlyGlyAspArgGlyArgGlyLysProGlyGlyMetArg---GlyGlyArgGlyGly 259
 DB 1441 CCGGGA-----GGGGGACAGGTGGCTCTCACTGGGGGGTAACTACGGG 1485
 QY 260 LeuMetAspArgGlyGlyProGlyGlyMetPheArgGly-----GlyArgGlyGly 276
 DB 1486 GATGATCGTGTGTGGCAGAGGAGGTATGATCGAGGCGCTACCGGGGCGGGGCGG 1545
 QY 277 AspArgGlyGlyPheArgGlyGlyArg---GlyMetAspArgGlyGlyPheGlyGlyGly 295
 DB 1546 GACCGTGGAGGCTTCGAGGGGCGCGGCGTGTGGGACAGAGTGTGCTTT-----1596
 QY 296 ArgArgGlyGlyProGlyGlyProProGlyProLeuMetGluGlnMetGlyGlyArgArg 315

DB 1596 -----1596
 QY 316 GlyGlyArgGlyGlyProGlyLysMetAsp---LysGlyGluHisArgGlnLysArg 334
 DB 1597 -----GGCCCTGGCAAGATGATTCAGAGGGTGAAGACAGACAGATCCGAG 1644
 QY 335 AspArgProGlyTr 338
 DB 1645 GAGAGGCGCTAT 1656

RESULT 12

US-09-919-039-322
 ; Sequence 322, Application US/09919039
 ; Publication No. US20030108871A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Kaefer, Matthew R.
 ; TITLE OF INVENTION: GENES EXPRESSED IN TREATED HUMAN C3A LIVER CELL CULTURES
 ; FILE REFERENCE: PA-0035 US
 ; CURRENT APPLICATION NUMBER: US/09/919, 039
 ; PRIOR FILING DATE: 2002-09-09
 ; PRIOR APPLICATION NUMBER: 60/222,113
 ; PRIOR FILING DATE: 2000-07-28
 ; NUMBER OF SEQ ID NOS: 401
 ; SOFTWARE: PERL Program
 ; SEQ ID NO 322
 ; LENGTH: 1939
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: misc feature
 ; OTHER INFORMATION: Incyte ID No. US20030108871A1 478620.53
 US-09-919-039-322

Alignment Scores:

Pred. No.: 7.33e-66 Length: 1939
 Score: 878.00 Matches: 182
 Percent Similarity: 63.37% Conservative: 36
 Best Local Similarity: 52.91% Mismatches: 54
 Query Match: 45.68% Indels: 72
 DB: 10 Gaps: 13

US-10-791-017a-2_COPY_319_656 (1-338) x US-09-919-039-322 (1-1939)

QY 1 G1G1YARGLYGLYMETGLYSERIALGLYUARGLYGLYPHEANLYSPROGLYGLY 20
 DB 825 GAGGAGAGAGTGGCATGGCCGGAAGT---GACCGTGTGGCTTCATTAATTTGGTGGC 881
 QY 21 ProMetAspGluGlyProAspLeuAspLeuGlyProProValAspProAspGluAspSer 40
 DB 882 CCTCGGAGCCAGGATGATCAGTCATGAC-----TCCGAACAGATTAATTTCA 926
 QY 41 AspAspSerAlaIleTyValGlnGlyLeuAspAspSerValThrLeuAspAspLeuAla 60
 DB 927 GACAAACAACCATCTTGTGCAAGGCTGGTGAATGTTCATTAATTTGATGCTGCTGCT 986
 QY 61 AspPhePheLeuGlnCysGlyValAlaIleYmeCAsnLysArgThrGlyGlnProMetIle 80
 DB 987 GATTACTTCAAGCAGATGGTATTTATTAACAACAAGAAACGGACAGCCCATGATTT 1046
 QY 81 HisIleTyLeuAspLysGlnThrGlyLysProLysGlyAspAlaThrValSerTyGlu 100
 DB 1047 AATTGTGACACAGACAGGAACTGGCAGCTGAAGGAGAGGCAACGGCTCTTTGAT 1106
 QY 101 AspProProThrAlaLysAlaAlaValAlaGluTrpPheAspGlyLysAspPheGlnGlySer 120
 DB 1107 GACCCACCTTCACCTTAAGCAGCTATGACTGTTGATGATAAGAAATTTCCGGAAT 1166
 QY 121 LysLeuLysValSerLeuAlaArgLysLysProLysProLysSerMetCArgGlyLysLeu 140
 DB 1167 CCTATCAAGTCTCTATTGCTACTCGCCGGGACAGACTTTAAT---CGGGGTGGTGGC 1220
 QY 141 ProProArgGluGlyArgGlyMetProProProLeuArgGlyGlyProGlyGlyProGly 160

```

Db      1221  AATGTCGTGAGAGCCGAGG-----CGAGAGAGACCCATGGCCGTGGA 1265
Qy      161  GLYPARGLYGLYPROMETGLYARGMETGLYGLYGLYASPARGLYGLYPHEPRO 180
      1266  GGCTATGAGAGTGGT-----GGCAGTGGTGGTGGCCGAGAGGATTTCCC 1313
Qy      181  ProxArgLYProArgLYSerArgLYAnProSerGLYGLYGLYAnValGLINHsArg 200
      1314  AGTGGAGGT-----GTTGGCCGTGGAGAGACAGCAGCA 1346
Db      201  AAGLYASPTPGLINCYPProAnProGLYCYGLYAnGLInAnPhealATPArgYthr 220
      1347  GCTGTGACGTGAGAGTGTCTTAATCCCACTGTGAGAAATGAACTTCTTGAGAGAA 1406
Qy      221  GLUCYASnGLINCYGLYsAlaProLYsPProGLUGLYPheLeuProProPheProPro 240
      1407  GAATGCACACAGTGAAGGCCCTTAACCAAGATGCC----- 1442
Qy      241  ProGLYGLYASPARGLYARGLYGLYPRoGLYGLYMeLArg---GLYGLYARGLYGLY 259
      1443  CCAAGA-----GGGGACCAAGTGGCTCTCAACATGGGGGTTACTACGG 1487
Db      260  LeuMetAsPARGLYGLYPRoGLYGLYMeLArgLY-----GLYARGLYGLY 276
      1488  GATGATCGTCGTGGTGGCAAGAGGCTATGATGAGCGGCTAACCGGGCCGCGGGGG 1547
Qy      277  AsPARGLYGLYPheArgLYGLYArg---GLYMeLAsPARGLYGLYPheGLYGLYGLY 295
      1548  GACCGTGAAGGCTTCCAGAGGGGGCGGGGTGGGACAGAGTGGCTTT----- 1598
Qy      296  ArgArgGLYGLYPRoGLYGLYPRoGLYProLeuMetGLUGLInMetGLYGLYArgArg 315
      1598  ----- 1598
Db      316  GLYGLYARGLYGLYPRoGLYGLYMeLArg---LYGLYGLYGLYHsArgGLInGLYArgArg 334
      1599  -----GGCCCTGGCAAGATGATTCAGAGGGGTGAGCAGACAGAGATCGCAGG 1646
Qy      335  AsPARGLYProTYr 338
      1647  GAGAGGCCGTAT 1658
Db

RESULT 13
US-09-918-995-9557
; Sequence 9557, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 9557
; LENGTH: 568
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(568)
; OTHER INFORMATION: n = A,T,C or G
; US-09-918-995-9557

Alignment Scores:
Pred. No.: 1,73e-60 Length: 568
Score: 809.50 Matches: 155
Percent Similarity: 97.48% Conservative: 0
Best Local Similarity: 97.48% Mismatches: 2
Query Match: 42.12% Indels: 3

```

```

DB:      10 Gaps: 1
US-10-791-017a-2_copy_319_656 (1-338) x US-09-918-995-9557 (1-568)
Qy      1  GLYGLYARGLYGLYMeLGLYSerAlaGLYGLYArgGLYGLYPheAnLYsPProGLYGLY 20
      92  GGAGGAGCGCGGTGAATGGC---GCTGAGAGAGCGAGTGGCTTCAATAGCCGTGTGGA 148
Qy      21  ProMetAsPGLUGLYProAsPLeuAsPLeuGLYProProValAsPProAsPGLYAsPser 40
      149  CCCATGATGAAGAGACCAATCTTATCTTAAGCCCACTGTGATCCAGATGAAGACTCT 208
Db      41  AsPAsnSerAlaIleTYrValGLInGLYLeuAsnAsPSeValThrLeuAsPAsPLeuAla 60
      209  GACCAACAGTGCATTTATTTATTCACAGGATTTAAATGACAGGTGACTTAAATGATCTGGCA 268
Qy      61  AsPAsPLeuGLINCYGLYValValLYsMeLAsnLYsArgThrGLYInPProMetIle 80
      269  GACTTCTTTAAGAGAGTGGGGGTGTGTTAAGATGACACAGAGACTGGCAACCAATGATC 328
Qy      81  HsIleTYrLeuAsPLeuAsPLeuThrGLYLYsPProLYsGLYAsPAlaThrValIserTYrGLu 100
      329  CACATTTACCTGGACAGAAACAGAAAGCCCAAGAGGATGCCACAGTCTCTATGAA 388
Db      101  AsPProProThrAlaLYsAlaAlaLYsValGLUTrPheAsPGLYLYsAsPheGLInLYSer 120
      389  GACCCACCACTGCTCCAGAGCTGCCGTGGAATGGTTGATGGGAAAGATTTTCAAGGGAGC 448
Qy      121  LYsLeuLYsValIserLeuAlaArgLYsLYsPProProMetAsnSerMeLArgGLYGLYLeu 140
      449  AAACCTTAAGTCTCCCTTCTCTCGAGAAAGCCCTCCAAATGAAACGATATGCG-GGTGGGCTG 507
Qy      141  ProProArgGLUGLYARGLYMeLProProProLeuArgLYGLYPRoGLYGLYPro 159
      508  CACCCCGTGAAGGGCAGAGGCAATGCCACCACTTTCGTGAGGTCCAGAGGCCCC 564
Db

RESULT 14
US-09-925-301-669
; Sequence 669, Application US/09925301
; Patent No. US20020052308A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA106
; CURRENT APPLICATION NUMBER: US/09/925,301
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05882
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1694
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 669
; LENGTH: 545
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (8)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (11)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (13)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (58)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (337)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature

```

LOCATION: (453)
OTHER INFORMATION: n equals a,t,c,g, or c
US-09-925-301-669

Alignment Scores:
Pred. No.: 1,01e-52 Length: 545
Score: 718.50 Matches: 141
Percent Similarity: 69.57% Conservative: 3
Best Local Similarity: 68.12% Mismatches: 19
Query Match: 37.38% Indels: 45
DB: 9 Gaps: 5

US-10-791-017a-2_COPY_319_656 (1-338) x US-09-925-301-669 (1-545)

```

QY 131 ProProMetAnsbSerMetArgGlyLeuProProArg-GluGlyArgGlyMetProPr 150
DB 17 CCTCACTTAAAGGAGACAAAGCTGAGCTCCACCGGCGTCCGCTTACGAACT-- 74
QY 150 oProLeuArgGlyProGlyProGlyProGlyProGlyProMetGlyArgMetG1 170
DB 75 -----AGTGAATCCCCCGG-----CTGACGAAATTCGG 103
QY 170 yGlyArgGlyGlyAspArgGlyGlyPheProProArgGlyProArgGlySerArgGlyAs 190
DB 104 CACGAGA-----GATGAGAGAGCTTCCCTCCAAAGAGACCCCGGGGTTCCGAGGGA 157
QY 190 nProSerGlyGlyGlyValGlnHisArgGlyAspTrpGlnCysProAnPro-- 209
DB 158 CCCCTCTGGAGAGAGAAAGCTCCAGCACCGAGCTGAGACTGCGACGTGCTCCATCCTTC 217
QY 210 -----GlyCysGlyValGlnHisPheAl 217
DB 218 AATTGCGATTCTCTGCTGATGATTAATGATGACGAGGAGTGTGAGAAACAGAACTTCGC 277
QY 217 aTPArgTrpGlnCysArgGlnCysGlyValAProGlyProGluGlyPheLeuProPro 237
DB 278 CTGGAGAAACAGAGTGCACACGATGT----- 302
QY 237 oPheProProProGlyGlyAspArgGlyArgGlyGlyProGlyGlyMetArgGlyGlyAr 257
DB 303 -----GATATCTGTGCAAGAGTGTGCTGTGTGCATGACGAGGAGAG 346
QY 257 gGlyGlyLeuMetAspArgGlyGlyProGlyGlyMetPheArgGlyGlyArgGlyGlyAs 277
DB 347 AGGTGGCTTATGATGATGTGTGTGTCCGGTGAATTTCAAGAGTGTGCTGTGTGAGA 406
QY 277 PATGlyGlyPheArgGlyGlyArgGlyMetAspArgGlyGlyPheGlyGlyGlyArgAr 297
DB 407 CAGAGGTGGCTTCCGTGTGTGTGCGGCGGCGCATGAGACGAGTGTGTTGTGAGAGAGAG 466
QY 297 gGlyGlyProGlyGlyProProGlyProLeuMetGluGlnMetGlyGlyArgArgGlyG1 317
DB 467 AGGTGGCTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 525
QY 317 yArgGlyGlyProGlyGlyAs 323
DB 526 ACGTGGAGGACCTGGGAAA 544

```

RESULT 15

US-09-960-352-3851
Sequence 3851, Application US/09960352
Patent No. US20020137139A1
GENERAL INFORMATION:
APPLICANT: Warren, Wesley C.
APPLICANT: Tao, Nengbing
APPLICANT: Byatt, John C.
APPLICANT: Mathalagan, Nagappan
TITLE OF INVENTION: MUSCLE AND FAT DEPOSITION
FILE REFERENCE: 16511.006/37-21(10298)C
CURRENT APPLICATION NUMBER: US/09/960,352
CURRENT FILING DATE: 2001-09-24
NUMBER OF SEQ ID NOS: 15112

SEQ ID NO 3851
LENGTH: 361
TYPE: DNA
ORGANISM: Bos taurus
OTHER INFORMATION: Clone ID: 17-LIB3057-024-Q1-K1-E1
US-09-960-352-3851

Alignment Scores:
Pred. No.: 3.73e-42 Length: 361
Score: 593.00 Matches: 111
Percent Similarity: 93.39% Conservative: 2
Best Local Similarity: 91.74% Mismatches: 7
Query Match: 30.85% Indels: 1
DB: 9 Gaps: 0

US-10-791-017a-2_COPY_319_656 (1-338) x US-09-960-352-3851 (1-361)

```

QY 67 GlyValValysMetAnlyAspArgTrpGlyGlnProMetIleHisIleTyrlLeuAspLys 86
DB 1 GGAGTCTTAAAGATGAACAGAGAGACTGTACAAACCATGATCCATATTACTTGACAGG 60
QY 87 GluThrGlyLysProLysGlyAspAlaThrValSerTyrlGluAspProProThrAlaLys 106
DB 61 TAAACGAAAGCCCAAGCGCATGCTACGCTCTTAAGAAAGCCAGCACTGCCAA 120
QY 107 AlaAlaValGluTrpPheAspGlyLysAspPheGlnGlySerTyrlLeuValSerLeu 126
DB 121 GCTGCTGTCCAGGTGTTGATGGAAAGATTTCCAAAGGAGCAAACTTAAAGTCTCTT 180
QY 127 AlaArgLysLysProProMetAnsbSerMetArgGlyGlyLeuProProArgGluArg 146
DB 181 GCTGGAGAGAGCCGCCCAAGACAGATCGGAGAGGATGCGCGCGAGGAGG 240
QY 147 GlyMetProProProLeuArgGlyGlyProGlyGlyProGlyGlyProGlyGlyProMet 166
DB 241 GGATGCGCGCGCGCGCTCCAGAGAGTCCAGGGGCGCAGAGAGTCTGTGAGAGACCATG 300
QY 167 GlyArgMetGlyGlyArgGlyGlyAspArgGlyGlyPheProProArg-GlyProArgG1 186
DB 301 GGGCGATGGAGAGCGCGGAGAGAGACAGAACTGCTTCCAAATGAGGGGCTTGGCGG 360
QY 186 Y 186
DB 361 T 361

```

Search completed: February 21, 2005, 08:39:51
Job time : 1887.22 secs

This Page Blank (uspio)